STRUCTURAL CHARACTERIZATION OF ENZYMES INVOLVED IN BIOSYNTHESIS OF THIAMIN AND SALVAGE OF URIDINE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Debamita Paul

May 2010

STRUCTURAL CHARACTERIZATION OF ENZYMES INVOLVED IN BIOSYNTHESIS OF THIAMIN AND SALVAGE OF URIDINE

Debamita Paul, Ph. D.

Cornell University 2010

Macromolecular X-ray crystallography was used for structural characterization of enzymes involved in thiamin biosynthetic and uridine salvage pathways. The enzymes involved in major metabolic pathways are potential drug targets and understanding their structural features and active site details lead to development of effective inhibitors. THI6 is a bifunctional enzyme present in the eukaryotic thiamin biosynthetic pathway. The N-terminal domain of this enzyme catalyses the ligation of the thiamin thiazole and pyrimidine moieties to form thiamin phosphate and the C-terminal domain catalyzes the phosphorylation of 4-methyl-5-(hydroxyethyl)thiazole in a salvage pathway. In lower organisms, thiamin phosphate synthase and 4-methyl-5-(hydroxyethyl)thiazole kinase are separate gene products. Current work reports the first crystal structure of a eukaryotic THI6 along with several complexes that characterize the active sites responsible for the two chemical reactions. THI6 from *Candida glabrata* is a homehexamer in which the six protomers form a cage like structure. Two domains within a protomer interact via two loop regions not found in the bacterial enzymes. The structures of different protein-ligand complexes define the active sites of the two domains. Our structural studies reveal that the active sites of the two domains are 40 Å apart and are not connected by an obvious channel.

The present work also includes structural and biochemical characterization of THI1 from *Zea maize*, a thiazole synthase involved in eukaryotic thiamin biosynthetic pathway. The structural work was performed to investigate the role of the residue Val211 in this particular enzyme, as genetic mapping showed that Val211Met mutation resulted in a phenotype called Bladekiller 1. In this phenotype plants develop some unusual characteristics, such as poor development of leaf blades, premature termination of meristems *etc*. The three dimensional structure of the wild-type enzyme in a complex form with the product have been described. Biochemical assays reveal that the wild type enzyme purified with the bound metabolites but the mutant does not. The dynamic light scattering studies show that the wild type enzyme forms an octamer in solution state unlike the mutant enzyme.

Finally the three dimensional structure of uridine phosphorylase, a key enzyme in the pyrimidine salvage pathway was described with the formation of glycal in its active site. This enzyme catalyzes the reversible phosphorolysis of uridine to uracil and ribose 1-phosphate (or 2′-deoxyuridine to 2′-deoxyribose 1-phosphate). The reaction is believed to proceed via an oxocarbenium ion-like transition state. The reported structures include *Escherichia coli* uridine phosphorylase treated with 5-fluorouridine and sulfate and dimeric bovine uridine phosphorylase treated with 5-fluoro-2′-deoxyuridine or uridine, plus sulfate. In each case the electron density shows three separate species corresponding to the pyrimidine base, sulfate and a ribosyl species. NMR time course studies demonstrated that uridine phosphorylase can catalyze the hydrolysis of these fluorinated nucleosides in the absence of phosphate or sulfate. Crystallization with the hydrolysis products in the presence of sulfate is then proposed to result in the *anti*-elimination of water across the C1-C2 bond of UPase-bound ribose. In the structures of the glycal complexes the fluorouracil O2 atom is appropriately positioned to act as the general base required for this elimination

reaction. Crystals of bovine uridine phosphorylase treated with 2′-deoxyuridine and sulfate show intact nucleoside. This nucleoside is unactivated toward cleavage of the *N*-ribosyl bond in the absence of phosphate. These results add a previously-unencountered mechanistic motif to the body of information on glycal formation catalyzed by enzymes which catalyze the cleavage of glycosyl bonds.

BIOGRAPHICAL SKETCH

Debamita Paul, daughter of Dr. Goutam Paul and Mrs. Kakoli Paul, was born in Durgapur an industrial town in West Bengal, India. She spent part of her childhood in a hill-station called siliguri and finished her primary education in Kolkata. Debamita developed her interest in chemistry during the final years of high school and that motivated her to join Lady Brabourne College, Kolkata to pursue her Bachelor of Science Degree in Chemistry.

After completion of college education in 2002, she joined the Master of Science Program in the prestigious Indian Institute of Technology, Mumbai. The two years program included some really interesting and useful courses in physical, inorganic and organic chemistry. Debamita was also involved in several research projects which including synthesis and characterization of different Ruthenium complexes.

During her involvement in the projects Debamita gained good interest in research and that directed her towards joining the graduate school. The author joined the Department of Chemistry and Chemical Biology in Cornell University, Ithaca, NY in Fall 2004. Her interest in the structure and function of enzymes motivated her to join Professor Steven. E. Ealick's research laboratory. For the past five years of her stay in Cornell, she has studied enzymes involved in different biosynthetic and salvage pathways using macromolecular X-ray Crystallography.

Apart from her fascination in structural biology, Debamita is also interested in music. She has received vocational training in Indian Classical Music for a long period. She also likes dancing, traveling and reading in her leisure.

To my parents.

ACKNOWLEDGMENTS

I would like to specially thank my advisor, Professor Steven E. Ealick, for teaching me the principles of Macromolecular X-ray Crystallography and for allowing me to work on multiple exciting projects in his laboratory. The support, assistance and guidance I got from him were really helpful in accomplishing my goals. I am grateful to my advisor for helping me towards building an analytical and critical view point of my work in my graduate career. I would like to thank Professor Tadhg Begley for being a collaborator on the thiamin biosynthesis project and for serving on my committee. I thank Professors Brian Crane and for serving on my committee and providing insights into my research work. I want to thank Ms. Leslie Kinsland for her assistance in various ways specially for making the delicious cookies. I thank all the former and present members of Ealick lab for their wonderful support and technical help. I thank Cynthia Kinsland and staff of the protein production facility for the cloning the plasmids. I thank the Begley lab members, Abhishek Chatterjee and Dinuka Abeydeera for the fruitful collaboration and helpful biochemical insights. I thank John Woodward, in Prof. Michael Scanlon's lab, for the collaboration with the *Zea maize* thiazole synthase project. I thank the staff of NE-CAT and CHESS for their technical assistance with data collection. I thank all my friends at Cornell and outside for their moral support throughout and for making the journey so special. Finally I want to thank my family for being with me in all the challenging moments and giving me constant motivation and strong support.

TABLE OF CONTENTS

Biographical Sketch iii

Dedication iv

Acknowledgements v

Table of Contents vi

List of Figures viii

List of Tables x

List of Abbreviations xi

Chapter 1. Introduction 1

References 11

Chapter 2. The Crystal Structure of *Candida glabrata* THI6, a Bifunctional Enzyme Involved in Thiamin Biosynthesis of Eukaryotes

Section 2.1. Introduction 14

Section 2.2. Experimental Procedures 16

Section 2.3. Results 25

Section 2.4. Discussion 38

References 47

Chapter 3. Structural studies on *Zea maize* Thiazole Synthase : Role of Val211

Section 3.1. Introduction 52

Section 3.2. Experimental Procedures 56

Section 3.3. Results 64

Section 3.4. Discussion 75

Appendix 82

References 83

Chapter 4. Glycal Formation in Crystals on Uridine Phosphorylase

Section 4.1. Introduction 87

Section 4.2. Experimental Procedures 89

Section 4.3. Results 99

Section 4.3 Discussion 110

References 118

Chapter 5. Summary 125

References 130

LIST OF FIGURES

Figure 1.1. Thiamin biosynthesis in prokaryotes 5

Figure 1.2. Thiamin biosynthesis in eukaryotes 6

Figure 2.1. Chemical reaction catalyzed by CgTHI6 15

Figure 2.2. Overall structure of CgTHI6 26

Figure 2.3. Central cavity of CgTHI6 27

Figure 2.4. Interaction between two domains within CgTHI6 protomer 28

Figure 2.5. Active site of CgTHI6 33

Figure 2.6. Activity assays for CgTHI6 37

Figure 2.7. Comparison of CgTHI6 with its bacterial homologue 39

Figure 2.8. Comparison of dimeric subunit of CgTHI6 with other enzymes 42

Figure 3.1. Thiazole biosynthesis in prokaryotes 54

Figure 3.2. Thiazole biosynthesis in eukaryotes 55

Figure 3.3. Structure of ZmTHI1 66

Figure 3.4. Active site of ZmTHI1 67

Figure 3.5. HPLC analysis for bound metabolites purifying with ZmTHI1 69

Figure 3.6. Biochemical assays for reconstitution of enzyme activity 70

Figure 3.7. DLS studies on wild type and mutant ZmTHI1 72

Figure 3.8. HPLC analyses for bound metabolites purifying with ZmTHI1. 73

Figure 3.9. DLS studies on wild type and mutant ZmTHI1 74

Figure 3.10 Comparison of ZmTHI1, AtTHI1 and ScTHI4 76

Figure 3.11 Val211 in the interface between four monomers 79

Figure 4.1. Electron density for EcUP and bUP complexes 102

Figure 4.2. Structures of EcUP and bUP 103

Figure 4.3. Active sites of EcUP and bUP 107

Figure 4.4. Hydrolysis of Urd in the presence of UP 110

Figure 4.5. Stereodiagram showing the conformational change upon ligand binding in bUP 112

Figure 4.6. Comparison of EcUP and bUP 113

LIST OF TABLES

Table 2.1. CgTHI6 data collection and processing statistics 20

Table 2.2. CgTHI6 refinement statistics 23

Table 2.3. Hydrogen bonds between the N-terminal and C-terminal domains of CgTHI6 29

Table 2.4. Comparison of the active site residues of BsTPS and CgTHI6 40

Table 2.5. Comparison of the active site residues of BsThiK and CgTHI6 44

Table 3.1. ZmTHI1 data collection statistics 59

Table 3.2. ZmTHI1 refinement statistics 61

Table 3.3. Dynamic light scattering 71

Table 3.4. Comparison of the active site residues of ZmTHI1, AtTHI1 and ScTHI4 77

Supplemental Figure 2.1. Primers used in cloning of ZmTHI1 82

Table 4.1. Data collection statistics for UP crystals 94

Table 4.2. Refinement statistics of UP crystals 96

Table 4.3. Analytical ultracentrifugation 108

LIST OF ABBREVIATIONS

Candida glabrata THI6*,* CgTHI6; Tobacco etch virus, TEV; β-mercaptoethanol, BME; noncrystallographic symmetry, NCS; Dithiothreitol, DTT; thiamin phosphate synthase, TPS; *Bacillus subtilis* thiamin phosphate synthase, BsTPS; *Pyrococcus furiosus* thiamin phosphate synthase, PfTPS*;* 4-methyl-5-(hydroxyethyl)thiazole kinase, ThiK; *Bacillus subtilis* thiazole kinase, BsThiK; thiamin monophosphate, ThMP; thiamin pyrophosphate, ThDP; pyrophosphate, PP; 4-amino-5-(hydroxymethyl)-2-methylpyrimidine phosphate, HMP-P; 4-amino-5-(hydroxymethyl)-2-methylpyrimidine pyrophosphate, HMP-PP; 4-amino-5-(hydroxymethyl)-2-(trifluoromethyl)-pyrimidine pyrophosphate, CF_3HMP-PP; 4-methyl-5-(hydroxyethyl)thiazole alcohol, Thz; 4-methyl-5-(hydroxyethyl)thiazole phosphate, Thz-P; 2-carboxy-4-methyl-5-(hydroxyethyl)thiazole phosphate, CO_2-Thz-P; adenosine 5′-triphosphate, ATP; *β,γ*-methylene adenosine 5′-diphosphate, AMP-PCP; adenosine diphospho-5-(β-ethyl)-4-methyl-thiazole-2-carboxylic acid, ADT; 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, HEPES; polyethylene glycol400, PEG400; Uridine phosphorylase, UP; *Escherichia coli* uridine phosphorylase, EcUP; bovine uridine phosphorylase, bUP; human uridine phosphorylase, hUP; purine nucleoside phosphorylase, PNP; nucleoside phosphorylase, NP; uracil, Ura; 5-florouracil, FUra; uridine, Urd; 5-fluorouridine, FUrd; 5-fluoro-2′-deoxyuridine, FdUrd; 2′-deoxyuridine, dUrd; selenomethionine, SeMet; isopropyl-1-β-D-galactopyranoside, IPTG; 2-(N-morpholino)ethanesulfonic acid, MES; tris(hydroxymethyl)aminomethane, Tris; PEG 5K monomethylether, PEG 5 K MME; Advanced Photon Source, APS; single wavelength anomalous diffraction, SAD; thiamin pyrophosphate, TPP; bladekiller 1, blk1; *Saccharomyces cerevisiae* THI4, ScTHI4; 1-deoxy-D-xylulose-5-phosphate, DXP; 1-deoxy-D-xylulose-5-phosphate

synthase, DXP; nicotinamide adenine nucleotide, NAD; *Arabidopsis thaliana* THI1, AtTHI1; *Zea maize* THI1_2, ZmTHI1; dynamic light scattering, DLS.

CHAPTER 1

INTRODUCTION

Overview of Structural Biology. Proteins are macromolecules that play major role in every central process of life, including catalyzing an extraordinary range of chemical reactions, composing the structural support of the cell, controlling membrane permeability, regulating the concentration of metabolites, causing motion and controlling the gene function. The structural study of these macromolecules at atomic resolution provides key insight into the aspects of their function and leads to the understanding of the underlying mechanisms of action. The field of structural biology is dedicated to this interrelationship between the structure and function of these macromolecules. In recent years, the remarkable development in the field of structural biology is also showing the way for a deeper understanding of evolution and the complex biological systems that governs cellular function.

The physical interactions studied in the field of structural biology are diverse. These levels include understanding the intramolecular interactions in macromolecules, leading to the proper folding from a polypeptide chain to secondary, tertiary and quaternary structures. Structural data also illuminates the interaction between the macromolecules and small molecules, including enzyme-substrate, transporter-metabolite and receptor-hormone interactions. Current progress in the field has extended the level of study of macromolecular assemblies including the study of protein-protein, protein-nucleic acid complexes.

One of the major goals of structural biology is to understand the catalytic reactions performed by enzymes. The structural details at atomic level provide information of the enzyme active site along with the identification of the active site residues. The three dimensional structure of enzymes with substrate, product and intermediate analogues provide the snapshots along the reaction coordinate. Structural

studies combined with biochemical assays and biophysical techniques can lead to the complete elucidation of the underlying mechanism in enzyme catalysis. The idea of transition-state formation by enzymes, was first recognized by Pauling in the year 1946 (*1*). 20 years after that Jencks established a fundamental connection between binding energy and enzyme catalysis (*2*). The success in the field of enzymology has been astounding in the past years. With the sufficient understanding of enzyme active sites and bioorganic chemistry the mechanism of numerous chemical transformations have been unraveled.

Another important application of structural biology is identification of several enzymes as drug target allowing for rational drug design. The three dimensional structure of a drug target and its interaction with the small molecule is used as a guide towards the drug-designing followed by optimization and virtual screening (3, 4). Current drug targets include not only enzymes, but also receptors, ion-channels, transporters and nucleic acids. A classic example was designing drugs against HIV protease, a key enzyme in the life cycle of HIV (4). Another example is discovery of Thymitaq®, a new treatment for liver cancer (5).

X-ray Crystallography as a Tool for Characterizing Enzymes. X-ray crystallography is considered to be one of the most powerful techniques in studying the three dimensional structure of a wide variety of macromolecules at atomic level. The field has been evolved in the last fifty years with the development of suitable instrumentations and computer programs. The first diffraction image from a protein crystal of pepsin was recorded in 1934 (6) and the protein structure solved was myoglobin in 1957, at 6 Å (7). The present statistics shows that to date about 64,500 protein structures have been solved including approximately 10,000 unique structures. The significant increase in the rate of solving structures is due to major advancement in the field of computer science, physics, material science and molecular and cell

biology. In the recent years the field of X-ray crystallography has also been playing a vital role in the work of Nobel laureates. Recenyly in 2009, the Nobel prize in chemistry was awarded to Ada E. Yonath, Thomas A. Steitz and Venkatraman Ramakrishnan for the detailed mapping of the ribosome using X-ray crystallographic techniques (8). The 2007 Nobel prize was rewarded to Roger Kornberg for his work on "the molecular basis of eukaryotic transcription", which included the X-ray structure of RNA polymerase II in complex with different transcription factors (9) . In 2003, the Nobel prize was given to Roderick Mackinnon and Peter Agre for the X-ray crystals of potassium and water channels respectively (10, 11).

The use of synchrotron radiation as the source for the experiments in X-ray crystallography provides highly coherent brilliance with tunable frequency. The application of new technology in building detectors with fast read out time, low noise and good spatial resolution has also improved the quality of experiments (*12*). The vast collection of computer programs used in data processing, phasing of anomalous data, molecular replacement, model building and refinement has been developing sharply with the availability for academic use.

Current Challenges in X-ray Crystallography. However there are few key challenges remaining in the field. The main one being production of good diffraction quality crystals, which is an intrinsic property of the protein. This problem has been approached in multiple ways such as, extensive high throughput screening, mutation of putative surface residues to reduce entropic effects, truncation of the terminal domains, chemical crosslinking *etc*. The most effective method is screening of different orthologues of a protein. Since the crystal packing is highly affected by the surface residues, hardly conserved among the orthologues, screening of different orthologues increase the probability of finding good crystals. The other limitation is dealing with small crystals for the difficulty in handling and their susceptibility to

radiation damage. This problem is being addressed by development of microbeam and robotics for a better mounting and data collection with the micro crystals.

Summary of Thesis Work. In the current study the macromolecular X-ray crystallography has been used as the tool for determination of three dimensional structures of enzymes involved in thiamin biosynthetic and pyrimidine salvage pathways. The enzymes involved in cofactor biosynthetic pathways have been of great interest due to their application as suitable drug targets, the novel chemistry involved in the catalytic reactions. The comparison of biosynthetic pathways existing in prokaryotes and eukaryotes also provide with the evolutionary implication. In addition to the biosynthetic pathways there are some energy efficient salvage pathways where the preformed nucleotides are metabolized into precursors for the synthesis of other nucleotides. The roles of the purine and pyrimidine nucleotide metabolic pathways, in diseases like cancer and viral infection have indicated many of the involved enzymes as suitable drug target, hence triggering their structural study.

Thiamin (Vitamin B_1), an essential molecule for all living systems serves as an important cofactor for enzymes involved in carbohydrate and amino acid metabolism (*13*). Bacteria, plant and fungi can synthesize thiamin whereas most other eukaryotes rely on nutritional thiamin intake. Thiamin biosynthesis has been the focus of study for several years due to various reasons (*14*). The biosynthesis involves some unusually unprecedented chemistry hence important from the mechanistic point of view. Morover, the studies on biosynthesis can be used for commercial production of thiamin and its components by fermentation. The biosynthetic process is very well characterized in prokaryotes both structurally and mechanistically (Figure 1.1). On the other hand the eukaryotic biosynthetic pathway is still in an emerging state (Figure 1.2).

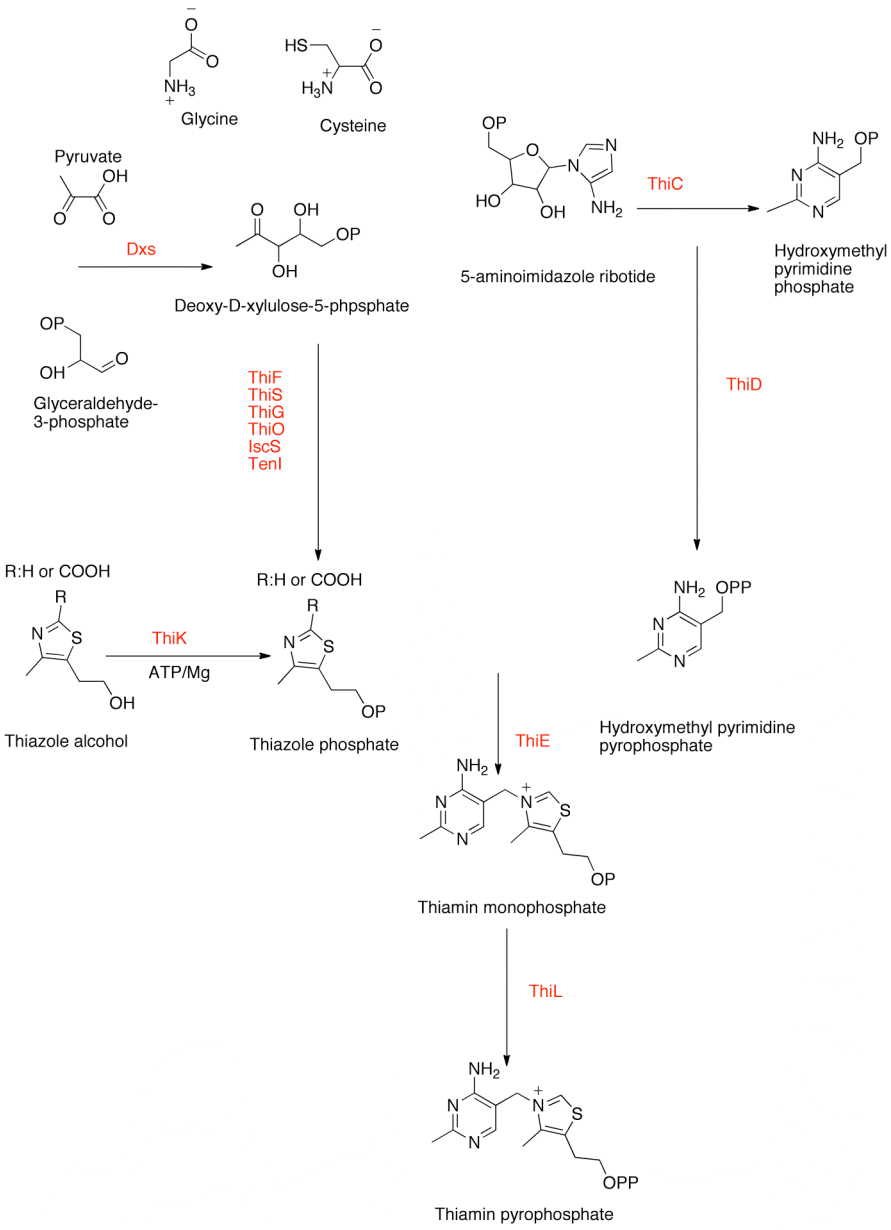


Figure 1.1. Thiamin biosynthesis in prokaryotes.

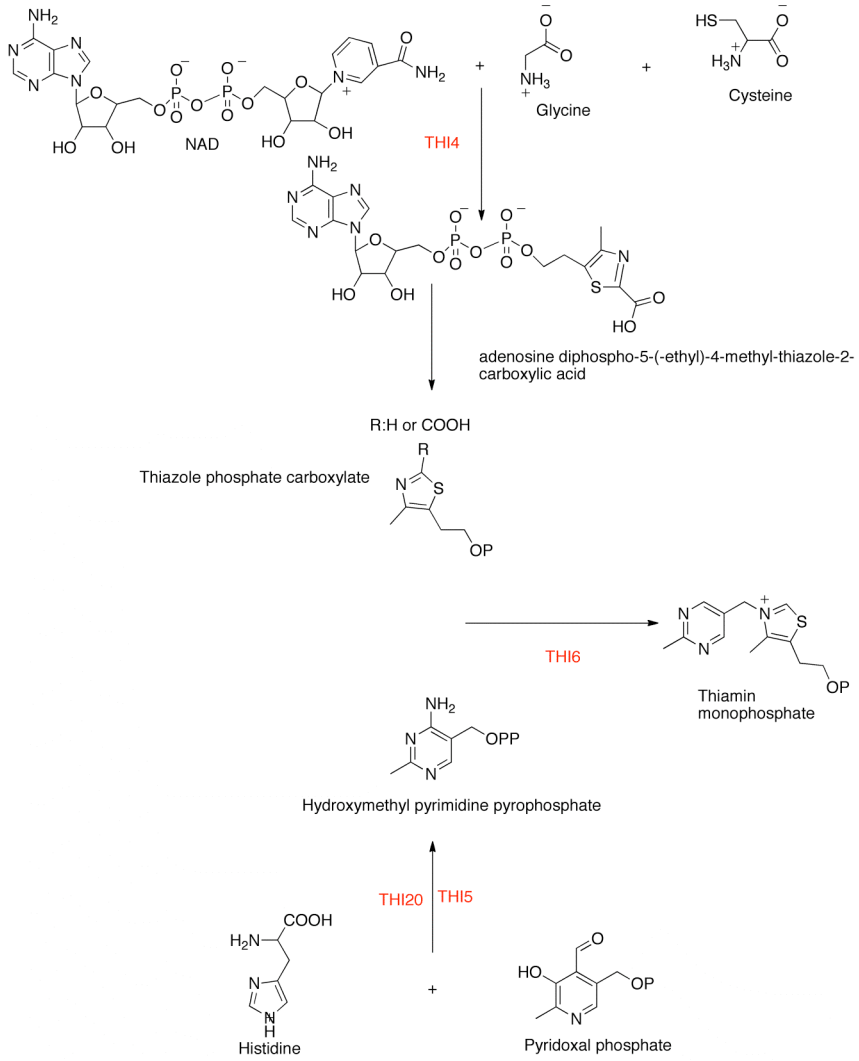


Figure 1.2. Thiamin biosynthesis in eukaryotes.

In Chapter 2, structural characterization of the bifuctional enzyme THI6, involved in thiamin biosynthesis of eukaryotes has been described. THI6 is a bifunctional enzyme found in the thiamin biosynthetic pathway of eukaryotes. The N-terminal domain of THI6 catalyses the ligation of the thiamin thiazole and pyrimidine moieties to form thiamin phosphate and the C-terminal domain catalyzes the phosphorylation of 4-methyl-5-(hydroxyethyl)thiazole in a salvage pathway. In prokaryotes, thiamin phosphate synthase and 4-methyl-5-(hydroxyethyl)thiazole kinase are separate gene products. Here we report the first crystal structure of a eukaryotic THI6 along with several complexes that characterize the active sites responsible for the two chemical reactions. THI6 from *Candida glabrata* is a homehexamer in which the six protomers form a cage like structure. Each protomer is composed of two domains, which are structurally homologous to their monofunctional bacterial counterparts. Two loop regions not found in the bacterial enzymes provide interactions between the two domains. The structures of different protein-ligand complexes define the thiazole and ATP binding sites of the 4-methyl-5-(hydroxyethyl)thiazole kinase domain, and the thiazole phosphate and 4-amino-5-(hydroxymethyl)-2-methylpyrimidine pyrophosphate binding sites of the thiamin phosphate synthase domain. Our structural studies reveal that the active sites of the two domains are 40 Å apart and are not connected by an obvious channel. Biochemical studies show that both 4-methyl-5-(hydroxyethyl)thiazole phosphate and the 2-carboxy-4-methyl-5-(hydroxyethyl)thiazole phosphate are substrates for THI6; however, adenosine diphospho-5-(β-ethyl)-4-methyl-thiazole-2-carboxylic acid, the product of THI4 is not a substrate for THI6. This suggests that unidentified enzyme may be necessary to produce the substrate for THI6 from the THI4 product.

Chapter 3 describes the structural and biochemical characterization of THI1 from *Zea maize*, a thiazole synthase involved in eukaryotic thiamin biosynthetic

pathway. The structural work was pursued to investigate the role of the residue Val211 in this particular enzyme, as genetic mapping showed that Val211Met mutation resulted in a phenotype called Bladekiller 1, in which maize plants develop some unusual characteristics, such as poor development of leaf blades, premature termination of meristems *etc*. The three dimensional structure of the wild-type enzyme in a complex form with the product have been described. Biochemical assays reveal that the wild type enzyme purified with the bound metabolites but the mutant does not. The dynamic light scattering studies show that the wild type enzyme forms an octamer in solution state unlike the mutant enzyme. Three additional mutants have been designed where the Val211 has been replaced by Ala (small side-chain), Leu (medium side-chain) and Phe (large side-chain) residues, to support our hypoyhesis that replacement of Val211 by a larger Met residue causes hindrance in thiazole biosynthesis and octamer formation. The crystal structure of V211A shows that the octameric state is retained by that mutant whereas some effects of the mutation are seen in thiazole biosynthesis.

Uridine phosphorylase is a crucial enzyme in the pyrimidine salvage pathway, which provides an alternative to the *de novo* biosynthetic pathway. This enzyme catalyzes the reversible phosphorolysis of uridine and 2′-deoxyuridine, as well as their analogues to respective nucleobases and ribose 1-phosphate. This enzyme plays a vital role in the homeostatic regulation of both intracellular and plasma uridine concentrations. It has been shown that the toxicity of antitumor (5-fluorouracil) (*15*) and anti-AIDs (3'-azido-3'-deoxythymidine) (*16*) drugs can be antagonized by administration of exogenous uridine. High doses of uridine are required to produce this therapeutic effect due to poor oral bioavailability and short half life of plasma uridine (two minutes) (*17*). However such large doses of uridine are not well tolerated and produce toxic side effects like phlebitis, pyrogenic reactions and diarrhea,

resulting form the accumulation of uridine catabolites. Inhibition of uridine phosphorylase activity can maintain an adequate concentration of uridine in the blood to alleviate the toxic effects of drugs (*18*) This modulation of uridine catabolism by inhibitors has been successfully used to attain the effect of substantial doses of uridine without the toxic side effects. The manipulation of uridine pool using uridine phosphorylase inhibition is also extended to the treatment of other pathological and physiological disorders like CNS disorders, sleep promotion, muscle performance, liver disease, cardiac damage and hereditary orotic aceduria. Hence the structural information about mammanlian uridine phosphorylase is immensely important for the drug designing viewpoint.

Chapter 4 describes the trapping of glycal in the active sites of bacterial (*Escherichia coli*) and mammalian (*Bos taurus*) uridine phosphorylase. Here we have reported the structure of hexameric *Escherichia coli* uridine phosphorylases treated with 5-fluorouridine and sulfate and dimeric bovine uridine phosphorylase treated with 5-fluoro-2′-deoxyuridine. In each case the electron density shows three separate species corresponding to the pyrimidine base, sulfate and a ribosyl species. NMR time course studies demonstrated that uridine phosphorylase can catalyze the hydrolysis of these fluorinated nucleosides in the absence of phosphate or sulfate. Crystallization with the hydrolysis products in the presence of sulfate is then proposed to result in the *anti*-elimination of water across the C1-C2 bond of UPase-bound ribose. In the structures of the glycal complexes the fluorouracil O2 atom is appropriately positioned to act as the general base required for this elimination reaction. Crystals of bovine uridine phosphorylase treated with 2′-deoxyuridine and sulfate show intact nucleoside. This nucleoside is unactivated toward cleavage of the *N*-ribosyl bond in the absence of phosphate.

These results add a previously-unencountered mechanistic motif to the body of information on glycal formation catalyzed by enzymes which catalyze the cleavage of glycosyl bonds.

REFERENCES

1. Pauling, L. (1956) The future of enzyme research, *Henry Ford Hosp Med Bull 4*, 1-4.
2. Page, M. I., and Jencks, W. P. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect, *Proc.Natl.Acad.Sci. U S A 68*, 1678-1683.
3. Blundell, T. L., Sibanda, B. L., Montalvao, R. W., Brewerton, S., Chelliah, V., Worth, C. L., Harmer, N. J., Davies, O., and Burke, D. Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery, *Philos Trans R Soc Lond B Biol Sci 361*, 413-423.
4. Tong, L., Pav, S., Pargellis, C., Do, F., Lamarre, D., and Anderson, P. C. Crystal structure of human immunodeficiency virus (HIV) type 2 protease in complex with a reduced amide inhibitor and comparison with HIV-1 protease structures, *Proc.Natl.Acad.Sci. U S A 90*, 8387-8391.
5. Tong, Y., Liu-Chen, X., Ercikan-Abali, E. A., Capiaux, G. M., Zhao, S. C., Banerjee, D., and Bertino, J. R. Isolation and characterization of thymitaq (AG337) and 5-fluoro-2-deoxyuridylate-resistant mutants of human thymidylate synthase from ethyl methanesulfonate-exposed human sarcoma HT1080 cells, *J. Biol.Chem. 273*, 11611-11618.
6. Bernal, J. D., and Crowfoot, D. (1934) X-ray photographs of crystalline pepsin, *Nature 133*, 794-795.
7. Kendrew, J. C. (1959) Structure and function in myoglobin and other proteins, *Fed.Proc. 18*, 740-751.
8. Sprinzl, M., and Erdmann, V. A. Protein biosynthesis on ribosomes in molecular resolution: Nobel Prize for chemistry 2009 goes to three chemical

biologists, *Chembiochem. 10*, 2851-2853.

9. Kornberg, R. D. The molecular basis of eukaryotic transcription, *Proc.Natl.Acad. Sci. U S A 104*, 12955-12961.
10. MacKinnon, R. Potassium channels, *FEBS.Lett. 555*, 62-65.
11. Agre, P., and Kozono, D. Aquaporin water channels: molecular mechanisms for human diseases, *FEBS.Lett. 555*, 72-78.
12. Walter, R. L., Thiel, D. J., Barna, S. L., Tate, M. W., Wall, M. E., Eikenberry, E. F., Gruner, S. M., and Ealick, S. E. High-resolution macromolecular structure determination using CCD detectors and synchrotron radiation, *Structure 3*, 835-844.
13. Harper, C. (2006) Thiamine (vitamin B1) deficiency and associated brain damage is still common throughout the world and prevention is simple and safe!, *Eur. J. Neurol. 13*, 1078-1082.
14. Begley, T. P., Downs, D. M., Ealick, S. E., McLafferty, F. W., Van Loon, A. P., Taylor, S., Campobasso, N., Chiu, H. J., Kinsland, C., Reddick, J. J., and Xi, J. (1999) Thiamin biosynthesis in prokaryotes, *Arch. Microbiol. 171*, 293-300.
15. Ashour, O. M., Naguib, F. N., Panzica, R. P., Al Safarjalani, O. N., and el Kouni, M. H. Modulation of 5-fluorouracil host toxicity by 5-(benzyloxybenzyl)barbituric acid acyclonucleoside, a uridine phosphorylase inhibitor, and 2',3',5'-tri-O-acetyluridine, a prodrug of uridine, *Biochem. Pharmacol. 60*, 427-431.
16. Sommadossi, J. P., Carlisle, R., Schinazi, R. F., and Zhou, Z. Uridine reverses the toxicity of 3'-azido-3'-deoxythymidine in normal human granulocyte-macrophage progenitor cells in vitro without impairment of antiretroviral activity, *Antimicrob. Agents Chemother 32*, 997-1001.

17. Ashour, O. M., Naguib, F. N., and el Kouni, M. H. 5-(m-Benzyloxybenzyl)barbituric acid acyclonucleoside, a uridine phosphorylase inhibitor, and 2',3',5'-tri-O-acetyluridine, a prodrug of uridine, as modulators of plasma uridine concentration. Implications for chemotherapy, *Biochem. Pharmacol. 51*, 1601-1611.

18. Al Safarjalani, O. N., Zhou, X. J., Naguib, F. N., Shi, J., Schinazi, R. F., and el Kouni, M. H. Enhancement of the bioavailability of oral uridine by coadministration of 5-(phenylthio)acyclouridine, a uridine phosphorylase inhibitor: implications for uridine rescue regimens in chemotherapy, *Cancer Chemother Pharmacol. 48*, 389-397.

CHAPTER 2

THE CRYSTAL STRUCTURE OF *CANDIDA GLABRATA* THI6, A BIFUNCTIONAL ENZYME INVOLVED IN THIAMIN BIOSYNTHETIC PATHWAY OF EUKARYOTES

Section 2.1 Introduction

Thiamin (vitamin B_1) is an essential component of all living systems. The active form of thiamin, thiamin pyrophosphate (ThDP), acts as a cofactor for several important enzymes in carbohydrate and amino acid metabolism (*1*). The mechanistic role of the cofactor is stabilization of an acyl carbanion intermediate (*2*). Thiamin biosynthetic pathways are found in prokaryotes and some eukaryotes (yeast, fungi and plants); however, vertebrates cannot synthesize thiamin. Therefore, thiamin is an essential component of the human diet with a daily requirement of 1.4 mg. Deficiency of thiamin leads to diseases such as beriberi and Wernicke-korsakoff syndrome (*3*). The absence of thiamin biosynthetic enzymes in mammals makes them potential targets for new antimicrobial and antifungal drug design.

Thiamin is composed of a thiazole and a pyrimidine moieties, which are synthesized separately and joined to form thiamin phosphate (ThMP). Prokaryotes and eukaryotes use different strategies for the production of thiamin. The detail of thiamin biosynthesis is well characterized both structurally and mechanistically in prokaryotes (*4*). Thiazole formation in bacteria requires six gene products and involves the oxidative condensation of 1-deoxy-D-xylulose-5-phosphate, glycine (in *Bacillus subtilis)* or tyrosine (in *Escherichia coli*) and cysteine (*5*). Formation of 4-amino-5-(hydroxymethyl)-2-methylpyrimidine phosphate (HMP-P) occurs by a complex rearrangement of 5-aminoimidazole ribonucleotide catalyzed by ThiC (*6*), a member of the radical SAM superfamily of enzymes. HMP-P is further phosphorylated by the enzyme ThiD to form 4-amino-5-(hydroxymethyl)-2-

methylpyrimidine pyrophosphate (HMP-PP) (*7*). The thiazole and pyrimidine modules are then coupled by the enzyme thiamine phosphate synthase (TPS) to form ThMP (Figure 2.1A) (*8*), which is finally phosphorylated to ThDP by the enzyme ThiL (*9*).

In addition to the biosynthetic pathway, many organisms contain salvage pathways. The enzyme 4-methyl-5-(hydroxyethyl)thiazole kinase (ThiK) catalyzes the adenosine 5′-triphosphate (ATP)-dependent phosphorylation of 4-methyl-5-(hydroxyethyl)thiazole (Thz) to form 4-methyl-5-(hydroxyethyl)thiazole phosphate (Thz-P) (Figure 2.1B) (*10*). ThiD is responsible for the phosphorylation of 4-amino-5-(hydroxymethyl)-2-methylpyrimidine (HMP) (*7*).

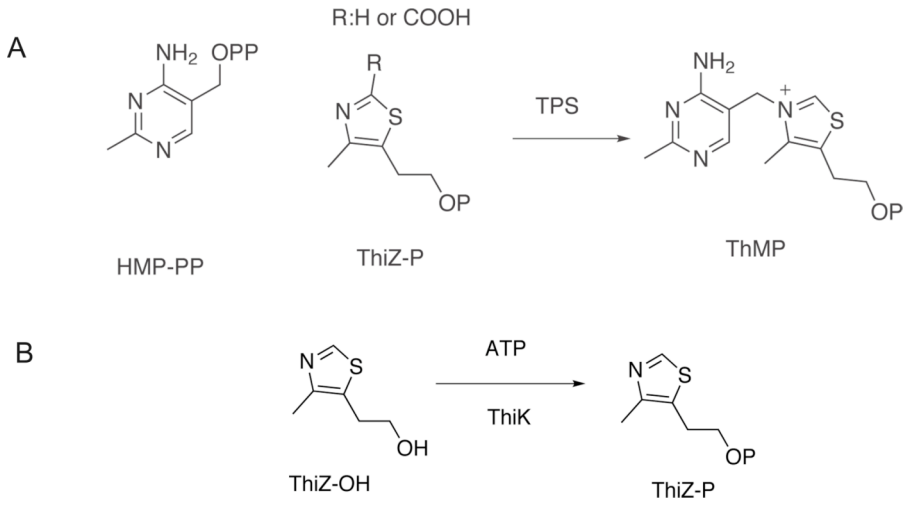


Figure 2.1. Chemical reactions catalyzed by CgTHI6. (A) Reaction catalyzed by the N-terminal thiamin phosphate synthase domain. (B) Reaction catalyzed by the C-terminal thazole kinase domain.

In contrast to the prokaryotic system, the mechanistic and structural understanding of thiamin biosynthesis is still at an early stage in the eukaryotes. Currently, only three enzymes are known to catalyze the biosynthesis of ThMP. THI4 converts nicotinamide adenine dinucleotide, glycine and a yet unknown sulfur source to form the adenosine diphospho-5-(β-ethyl)-4-methyl-thiazole-2-carboxylic acid (ADT) (*11, 12*). THI5 is responsible for HMP-PP formation; however, the mechanistic and structural details of this process remain unknown (*13*). The coupling of CO_2-Thz-P and HMP-PP, in eukaryotes, is catalyzed by THI6 (*14*). This bifunctional enzyme contains a TPS domain and a ThiK domain. Earlier studies on THI6 from *Saccharomyces cerevisiae* showed that the N-terminal domain is responsible for the TPS activity and the C-terminal domain is responsible for ThiK activity (*15*).

Here we report the crystal structures of unliganded *Candida glabrata* THI6 (CgTHI6) and of several complexes that map out the two active sites. These include ThMP; ThMP and pyrophosphate (PP); Thz and *β,γ*-methylene adenosine 5′-diphosphate (AMP-PCP); ThMP and AMP-PCP and 4-amino-5-(hydroxymethyl)-2-(trifluoromethyl)-pyrimidine pyrophosphate (CF_3HMP-PP) and CO_2-Thz-P. The THI6 structure reveals a cage-like homohexamer in which the two active sites are separated by about 40 Å. The two active sites are not connected by an obvious channel. We also report biochemical studies that show that the TPS domain accepts Thz-P as a substrate, but not ADT, the product of THI4.

Section 2.2 Experimental

Chemical Reagents. Kanamycin, ThMP, sodium pyrophosphate, Thz, AMP-PCP, ATP, magnesium chloride and the buffer components were all purchased from Sigma-Aldrich. Ni-NTA resin was obtained from Qiagen (Valencia, CA). HPLC grade solvents were obtained from Fisher Scientific. Thz-P and HMP-PP for the

activity assays were synthesized by Abhishek Chatterjee as previously described (*16*). The adenylated thiazole (Thz-ADP) was also described by Abhishek Chatterjee, by mixing, commercially available adenosine monophosphate morpholidate and Thz-P in dry pyridine. An overnight reaction at room temperature generated Thz-ADP which was purified by anion exchange chromatography (*17*). TEV protease was provided by Cornell Protein Production Facility.

Expression and Purification of CgTHI6. The CgTHI6 gene was cloned into a modified pET28 overexpression vector encoding an N-terminal hexahistidine tag with a TEV protease cleavage site. The plasmid CgTHI6.XF1 was transformed into *Escherichia coli* cell line BL21(DE3). A 10 mL overnight culture, grown in Luria-Bertani media at 37 ºC supplemented with 100 μg/mL ampicillin, was used to start a 1 L culture. Growth at 37 ºC was allowed until the OD_{600} reached 0.5 at which point the temperature was reduced to 15 ºC. After 45 min of cooling the culture was induced with 0.1 mM isopropyl-1-β -D thiogalactopyranoside (IPTG) and allowed to grow overnight at 15 ºC. Cells were harvested by centrifugation at 9000 *g* for 15 min at 4 ºC using an Avanti J-251 centrifuge (Beckman) and frozen at -20 ºC for later use.

The purification was carried out at 4 °C. The frozen cell pellet was thawed and resuspended in 50 mL of lysis buffer [50 mM Tris (pH 8.0), 300 mM NaCl, 10 mM imidazole and 3 mM β-mercaptoethanol (BME)] and lysed by sonication. Cellular debris was removed by centrifugation at 35000 *g* for 50 min at 4 ºC. The clarified lysate was loaded onto a Ni-NTA column (Qiagen) pre-equilibriated with 50 mL of the lysis buffer. The column was then washed with 100 mL of wash buffer [50 mM Tris (pH 8.0), 300 mM NaCl, 60 mM imidazole and 3 mM BME] to remove the nonspecific binding proteins. CgTHI6 was eluted with elution buffer [50 mM Tris (pH 8.0), 300 mM NaCl, 250 mM imidazole and 3 mM BME].

The elution was monitored by measuring protein concentration via Coomassie Plus Protein Assay Reagent (Pierce).

To remove the N-terminal hexahistidine tag, the eluted protein was incubated with TEV protease, in dialysis tubing (Pierce) with a 10 kDa molecular weight cutoff (MWCO) (Millipore) for 12 h at 4 °C, in a buffer containing, 50 mM Tris (pH 8.0), 150 mM NaCl, 1 mM DTT. The proteolysis was monitored by SDS-PAGE analysis. After complete proteolysis, the reaction mixture was reloaded onto a Ni-NTA column and the flow through was collected. The non-binding material was buffer exchanged into 10 mM Tris (pH 7.5), 50 mM NaCl. The protein was then concentrated to 8 mg/mL using a 30 kDa MCWO Amicon Ultracentrifugal filter (Millipore) and stored at -80 °C for future use. Protein concentration was determined by the Bradford method (*18*) with bovine serum albumin as standard. The purity of the protein was determined by Coomassie-stained SDS-PAGE analysis and was found to be greater that 95% (data not shown).

Crystallization of CgTHI6 in Unliganded and Complex Forms. The crystallization experiments were conducted at 22 ºC using hanging-drop vapor diffusion by combining 1 μL of protein solution and 1 μL of reservoir solution. Initial crystallization conditions of unliganded CgTHI6 were determined using sparse matrix screens, Crystal Screen 1 and 2 (Hampton research). The final optimized condition contained 0.1 M HEPES (pH 7.5), 0.2-0.25 M $MgCl_2$ and 25%-32% polyethylene glycol400 (PEG400). Rod shaped crystals, 200-300 μm long, appeared in 2-3 days.

Crystals of the THI6/ThMP complex were obtained by cocrystallization under similar conditions. The enzyme was incubated with 2.5 mM of ThMP at 4 ºC for 1 h prior to crystallization. The crystals grew after 2-3 days to the maximum size of 400 μm.

Additional complexes were prepared by soaking CgTHI6 crystals with the appropriate ligands. The THI6/ThMP/PP complex was prepared by transferring CgTHI6 crystals into a solution containing the mother liquor with 5 mM ThMP and 5 mM sodium pyrophosphate and soaking for 1 h at 22 ºC. The THI6/Thz/AMP-PCP was obtained by soaking CgTHI6 crystals in a solution containing the mother liquor with 20 mM Thz and 20 mM AMP-PCP for 12 h at 22 ºC. The THI6/ThMP/AMP-PCP complex was obtained by soaking the CgTHI6 crystals in mother liquor solution containing 20 mM each of ThMP, Thz and AMP-PCP for 2 h at 22 ºC. However Thz was not bound in the active site. The THI6/CF_3HMP-PP/CO_2-Thz-P complex was obtained by soaking CgTHI6 crystals in a solution containing 10 mM each of the ligands for 1 h at 22 ºC.

The high concentration of PEG400 in the mother liquor acted as a cryo-protectant. Crystals were flash frozen directly from the drop by plunging them into liquid nitrogen.

Data Collection and Processing. All data were collected at the NE-CAT beamlines 24-ID-E or 24 -ID-C at the Advanced Photon Source at Argonne National Laboratory using a Quantum 315 X-ray detector (Area Detector Systems Corporation). Crystals of CgTHI6 initially diffracted to ~ 3 Å but were very sensitive to radiation damage. As a result, a highly attenuated beam with 10 sec exposure time, 1º oscillation and 400 mm of crystal to detector distance were used to record the diffraction images. To obtain a complete dataset, multiple segments of data were collected by exposing the same rod-shaped crystal at different spots and merged together. The data were indexed, integrated and scaled using HKL2000 program suite (*19*). The data collection and processing statistics are shown in Table 2.1.

Table 2.1. Data collection statistics. Values for the highest resolution shell are given in parentheses.

	THI6	THI6/ThMP	THI6/ThMP–PP	THI6/ThiZOH/AMP-PCP	THI6/ThMP/AMP-PCP	THI6/CF_3H MP-PP/CO_2-Thz-P
Wavelength (Å)	0.9795	0.9795	0.9795	0.9795	0.9795	0.9795
Resolution (Å)	3.2	3.0	3.1	3.3	2.6	3.2
Space group	$P2_1$	$P2_1$	$P2_1$	$C2$	$C2$	$P2_1$
Unit Cell						
a (Å)	102.4	104.5	105.2	161.3	163.2	104.9
b (Å)	153.0	154.3	154.6	153.9	153.5	154.2
c (Å)	146.7	148.6	148.6	108.6	109.6	148.7
β (º)	102.3	102.5	102.1	117.6	117.9	102.1
Matthews coefficient	3.1	3.3	3.4	3.4	3.5	3.2
% solvent	60.1	63.3	63.7	64.1	64.7	62.3
mol/asym	6	6	6	3	3	6
Reflections	75076	273270	188896	163837	143976	159587
Unique reflections	38571	84611	72303	32447	69157	74009
Completeness (%)	93.8 (68.9)	93.2 (61.0)	86.6 (71.3)	91.5 (58.7)	96.5 (82.3)	98.5 (95.9)
Rsym *(%)	10.6 (32.8)	10.3 (26.9)	11.0 (34.8)	14.8 (39.7)	6.3 (35.5)	8.1 (42.6)
I/σ	13.7 (2.3)	9.5 (1.8)	8.5 (1.4)	7.8 (1.6)	11.6 (1.8)	12.8 (1.9)
Redundancy	2.1 (2.1)	3.2 (2.1)	2.6 (1.9)	5.0 (2.8)	2.1 (1.9)	2.2 (2.0)

*$R_{sym} = \Sigma\Sigma_i \left| I_i - \langle I \rangle \right| / \Sigma \langle I \rangle$, where <I> is the mean intensity of the N reflections with intensities I_i and common indices *h,k,l*.

Structure Determination and Refinement. Blast searches (*20*) showed that the N-terminal domain of CgTHI6 is about 30% identical to *Bacillus subtilis* TPS (BsTPS) and the C-terminal domain is about 30% identical to *Bacillus subtilis* ThiK (BsThiK). As a result, molecular replacement was used for the structure determination of unliganded CgTHI6. The Matthews number (*21*) suggested that the asymmetric unit contains six monomers with solvent content of 60.1%. A homology model containing three C-terminal domains of THI6 was generated using trimeric BsThiK (PDB entry 1C3Q) as a template. Only conserved side chains, based on sequence alignments of CgTHI6 with multiple bacterial thiazole kinases using ClustalW (*22*), were retained in the homology model. Molecular replacement using CNS (*23*) revealed the orientations and positions of two trimeric assemblies. After fixing the assemblies from the C-terminal domains, the six N-terminal domains were found by molecular replacement using MOLREP (*24*) with a monomer of BsTPS (PDB entry 2TPS) as the search model.

All model building was performed using the computer program COOT (*25*). The initial refinement was done using CNS and involved successive rounds of rigid body refinement, simulated annealing, temperature factor refinement, energy minimization and model rebuilding. Tight non-crystallographic symmetry restraints (NCS) were used throughout the refinement. The sixfold NCS averaged maps were generated using the RAVE suit of programs (*26*). NCS averaged electron density maps and averaged composite omit maps were used for all stages of model building. Final rounds of refinements were performed using the PHENIX suit of programs (*27*) with tight NCS restraints.

Structure determinations of the CgTHI6-ligand complexes were carried out using the unliganded structure as a starting point. The THI6/Thz/AMP-PCP complex crystallized in space group *C*2 with three CgTHI6 chains per asymmetric unit. The structure was readily determined by molecular replacement using half of a hexamer as the search model. Successive rounds of rigid body refinement, simulated annealing, temperature factor refinement and energy minimization were carried out, followed by model rebuilding. After few rounds of refinement the NCS averaged electron density map was used to position the ligands in the active site. PRODRG was used to generate the topology and parameter files of ligands, required for refinement (*28*). The final refinement statistics are summarized in Table 2.2.

Table 2.2. Refinement statistics.

	THI6	THI6/ThMP	THI6/ThMP/PP	THI6/Thz/AMP-PCP	THI6/ThMP/AMP-PCP	THI6/CF_3HMP-PP/CO_2-Thz-P
Resolution (Å)	3.2	3.0	3.1	3.3	2.6	3.2
No of protein atoms	22510	22678	22967	11429	11967	22584
No of ligand atoms	-	144	192	128	162	222
RMSD						
Bonds (Å)	0.009	0.008	0.007	0.006	0.008	0.010
Angles (deg)	1.165	1.012	1.115	1.084	1.120	1.227
R factor [a] (%)	21.38	21.25	20.50	20.44	21.23	20.1
R_{free} [b] (%)	24.48	24.18	25.10	24.29	24.40	23.2
Ramachandran plot						
Most favored (%)	86.1	88.4	85.2	81.6	88.0	85.7
Additional allowed (%)	12.3	10.0	12.8	15.8	11.2	12,2
Generously allowed (%)	1.6	1.4	1.7	2.3	0.8	1.8
Disallowed (%)	0.0	0.2	0.3	0.3	0.0	0.2

[a] $R\ factor = \Sigma_{hkl} \left| \left| F_{obs} \right| - k \left| F_{cal} \right| \right| / \Sigma_{hkl} \left| F_{obs} \right|$, where F_{obs} and F_{cal} are observed and calculated structure factors, respectively. In [b] R_{free} the sum is extended over a subset of reflections that were excluded from all stages of refinement.

Biochemical Assay of CgTHI6 Activity. The reaction mixtures of following composition were incubated at room temperature for 2 h. ThiK activity assay: 1.2 mM Thz, 2 mM ATP, 2 mM $MgSO_4$ and 33 μM purified enzyme in 50 mM Tris-HCl containing 150 mM NaCl and 2 mM dithiothreitol (DTT). TPS activity assay: 1.2 mM Thz, 2 mM ATP, 2 mM $MgSO_4$, 1 mM HMP-PP and 33 μM purified enzyme in 50 mM Tris-HCl containing 150 mM NaCl and 2 mM DTTs. ADT cleavage activity assay: 100 μm ADT, 2 mM $MgSO_4$ and 33 μM purified enzyme in 50 mM Tris-HCl containing 150 mM NaCl and 2 mM DTT. A reaction in which the enzymes was replaced by the reaction buffer served as a control.

HPLC Analysis of CgTHI6 Catalyzed Reactions. Analytical HPLC was carried out using a Supelco LC-18-T (150x4.6 mm, 3 μm ID) reverse phase column. HPLC analyses were performed in an Agilent 1100 instrument equipped with an inline diode-array detector. The enzyme was denatured by heat (100 °C, 40 s) and the reaction mixtures were rapidly cooled on ice. The precipitated protein was removed by centrifugation and the supernatant was filtered through a 10 kDa MWCO filter (Microcon YM-10, Millipore). 100 μL of the filtrate was analyzed by reversed phase HPLC. The following linear gradient was used at 1 mL/min flow rate (solvent A is water, solvent B is 100 mM KP_i, pH 6.6, solvent C is methanol): 0 min 100% B, 5 min 10% A 90% B, 8 min 25% A 60% B 15% C, 14 min 25% A 60% B 15% C, 19 min 30% A 40% B 30% C, 21 min 100% B, 30 min 100% B. The absorbance at 261 nm was monitored. Different compounds were identified in the chromatogram by comigration with authentic standards.

Figure Preparation. All figures were prepared using ClustalW (*22*), ESPript (*29*) or PYMOL (*30*).

Section 2.3 Results

Overall Structure of CgTHI6. The crystal structure shows that the enzyme is a hexamer composed of six identical protomers with 32 point symmetry (Figure 2.2A, B). The cage-like hexamer has the shape of a prolate spheroid with a diameter along the twofold axis of about 100 Å and a length along the threefold axis of about 150 Å. The inner cavity of the cage has a maximum diameter of about 55 Å (Figure 2.3). The presence of the cavity contributes to the high solvent content (~60%) of the crystals, and explains previous results from analytical size exclusion chromatography, which suggested THI6 was an octamer (*14*). The caps of the prolate spheroid are formed by trimeric assemblies of the C-terminal (ThiK) domains and the equator is formed by dimeric assemblies of the N-terminal (TPS) domains. The trimeric caps contact each other only through the TPS dimers; TPS dimers contact the trimers, but not each other.

An interesting aspect of the structure of this multidomain protein is the formation of the hexameric assembly by combination of the two trimeric C-terminal and three dimeric N-terminal domains. The two trimeric C-terminal domains form the top and bottom core of the structure, connected by three dimeric N-terminal subunit bridges.

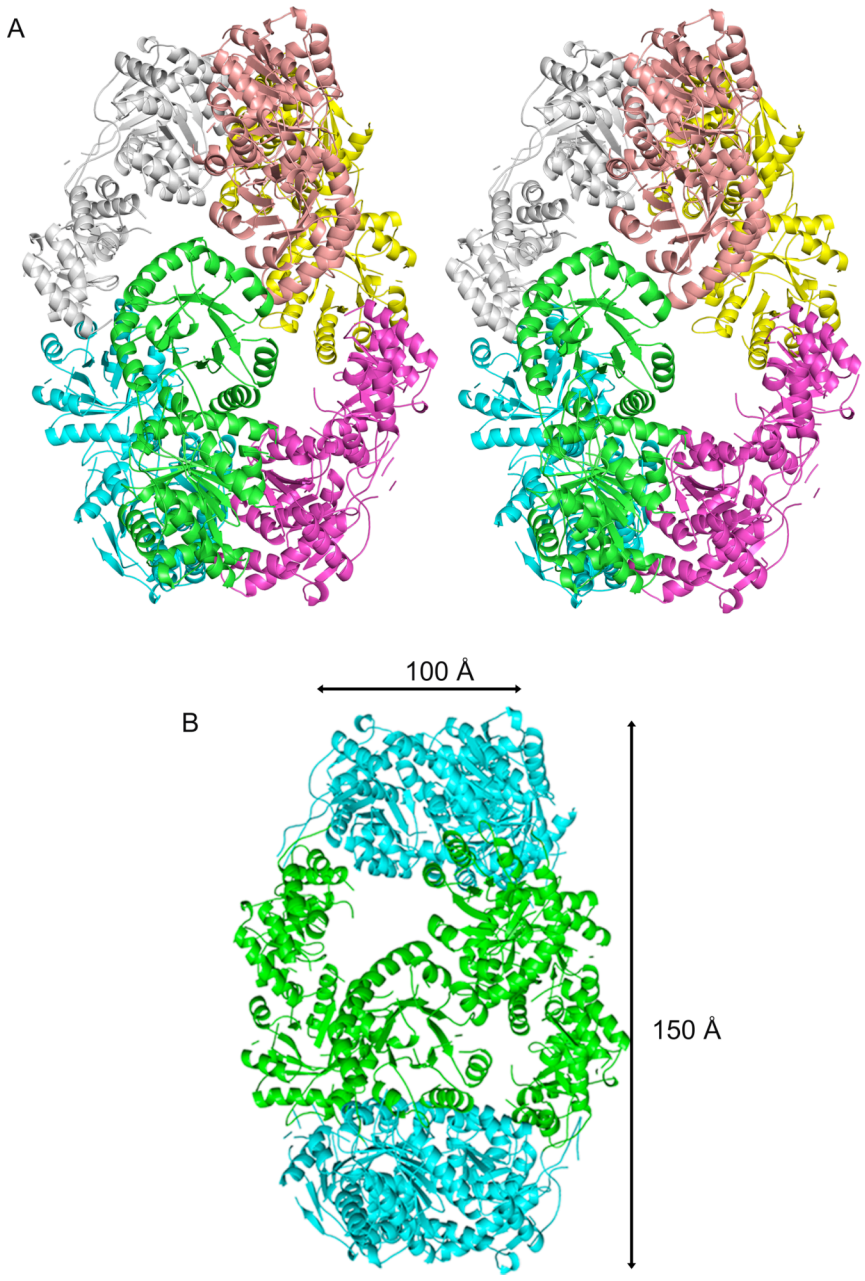


Figure 2.2. (A) Stereo diagram showing cartoon representation of the quaternary structure of CgTHI6 hexamer color coded by subunits. (B) Cartoon representation of the quaternary structure of CgTHI6 hexamer color coded by domains; N-terminal domains are colored in cyan and C-terminal domains are green.

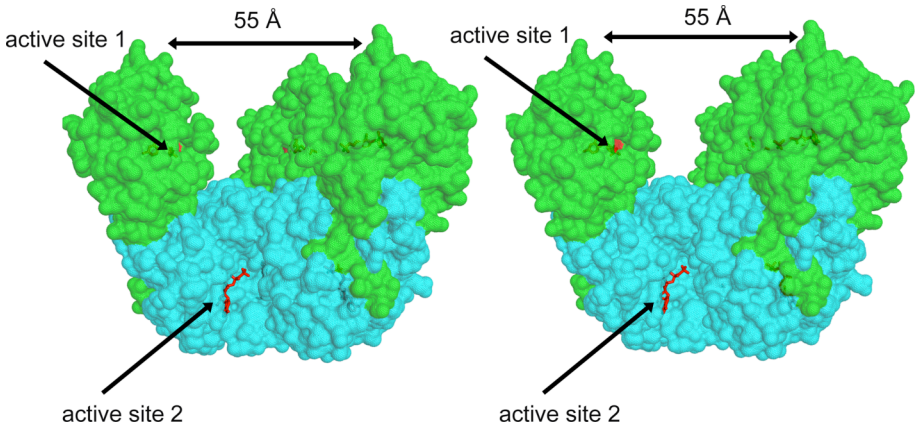


Figure 2.3. Stereo diagram showing the central cavity of CgTHI6, color coded by domain.

Cg*THI6 Protomeric Structure.* The protomer consists of two domains (Figure 2.2B). The N-terminal domain composed of the residues 1-236 and the C-terminal domain composed of residues 248-540. The domains are connected by a linker, consisting of 11 amino acid residues 237-247. The N-terminal domain has a $(\beta\alpha)_8$ barrel fold also known as TIM barrel fold. The structural fold is characterized by eight parallel β strands tilted at an angle of 45º from the barrel axis, surrounded by eight α helices. In addition to those eight α helices, flanking the barrel, there are two 3_{10} helices. The first one, α_0 forming a cap near the N-terminal end (residues 7-10) and the second one, $\alpha 8_0$ formed by residues 210-213, connecting β8 and α8. This helix is approximately tilted to angle of about 20º to the axis of the TIM barrel helices. The β strands are 4-6 residues long, whereas the α helices consist of 4-14 residues, α8 being the longest one. There are eight loops connecting the alternative helical and strand segments. The active site loop formed by residues 143-153 is disordered in the unliganded structure.

The C-terminal domain has an α/β fold. The centrally located, nine stranded mixed β sheet is flanked by 11 α helices. The β sheet has a topology: β′2↑ β′1↑ β′3↑

β′4↑ β′5↑ β′6↑ β′7↓ β′8↓ β′9↓. The helices α′3, α′4, α′5, α′6, and α′7 are packed on one side of the sheet being antiparallel to the strands of the sheets. The remaining six α helices form a bundle on the opposite side of the sheet. The strands contain 4-6 residues and the helices are made up of 6-14 residues. The C-terminal domains assemble to form a trimer in the top and bottom of the egg shaped structure.

Domain Interactions: Interface between N-terminal and C-terminal Domains. Within the protomer, 2630 Å^2 is buried at the interface of the TPS and ThiK domains (*31*). The interface is formed by several interactions (Figure 2.3, Table 2.3). The first interaction is between residues 193-203, a loop joining α7 and β8 of the TPS domain and a portion of the loop containing residues 432-438 of the ThiK domain.

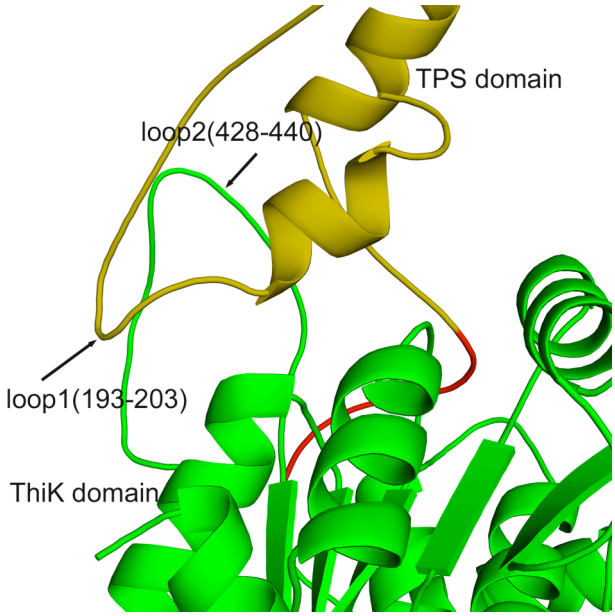


Figure 2.4. Ribbon diagram showing interactions between the two domains within CgTHI6 protomer. The N-terminal domain (residues 1-236) is colored in golden yellow, C-terminal (residues 248-540) in green and the linker region (residues 237-247) in red.

Table 2.3. Hydrogen bonds between the N-terminal and C-terminal domains.

N-terminal domain	C-terminal domain	H-bonding atoms
Ser202	Ser432	Oγ—N
Leu203	Ser434	O—N
Lys200	Ala435	O—N
Tyr234	Gly429	N—O
Arg189	Thr359	NH1—O

The second interaction occurs between residues 230-236 of the TPS domain and the loop, consisting of residues 428-432 from the ThiK domain. The third interaction is between part of α7 (residues 188-190) of the TPS domain and a portion of α′5 (residues 359-361) from the ThiK domain. Finally, the loop connecting α′10 and α′11 (residues 479-486) of the ThiK domain interacts with a portion of the linker (residues 238-241) between the TPS and ThiK domains.

Domain Interactions: Interface between N-terminal Domains. The six TPS domains of CgTHI6 have $(\beta\alpha)_8$ barrel folds and pair up to form three dimers. The individual domains are *cis* to each other with respect to the barrel axis. The total surface area buried at each dimeric interface is approximately 1500 $Å^2$ (*31*). The major interactions at dimer interfaces include helices α3 (residue 78-84) and α4 (residue 98-105). Helix α3 of one monomer is oriented almost perpendicular to the barrel axis of the other monomer, interacting with α3 and α4. A second interaction occurs between 49-52 from one domain and 94-97 from the other. The interface is mostly hydrophobic residues, with only one intermolecular hydrogen (Asp76 of one monomer and Arg77 of another).

Domain Interactions: Interface between C-terminal Domains. The ThiK domains group together to form two trimers, which are twisted by about 20º with respect to each other. The trimers are about 78 Å along the edge and about 45 Å thick. The total surface area buried between two ThiK domains is approximately 1600 $Å^2$ (*31*) and the interface is formed by thee main interactions. The first interaction occurs between part of α′2 (residues 276-279) from one domain and a portion of α′2 residues (273-275) from an adjacent domain. The second interaction occurs between a loop region from one domain (residues 297-303) and part of α′5 (347-350) form an adjacent domain. The side chain of Glu298 forms hydrogen bonds with the side chains of Thr348 and Arg350. The third interaction occurs between part of helix α′10

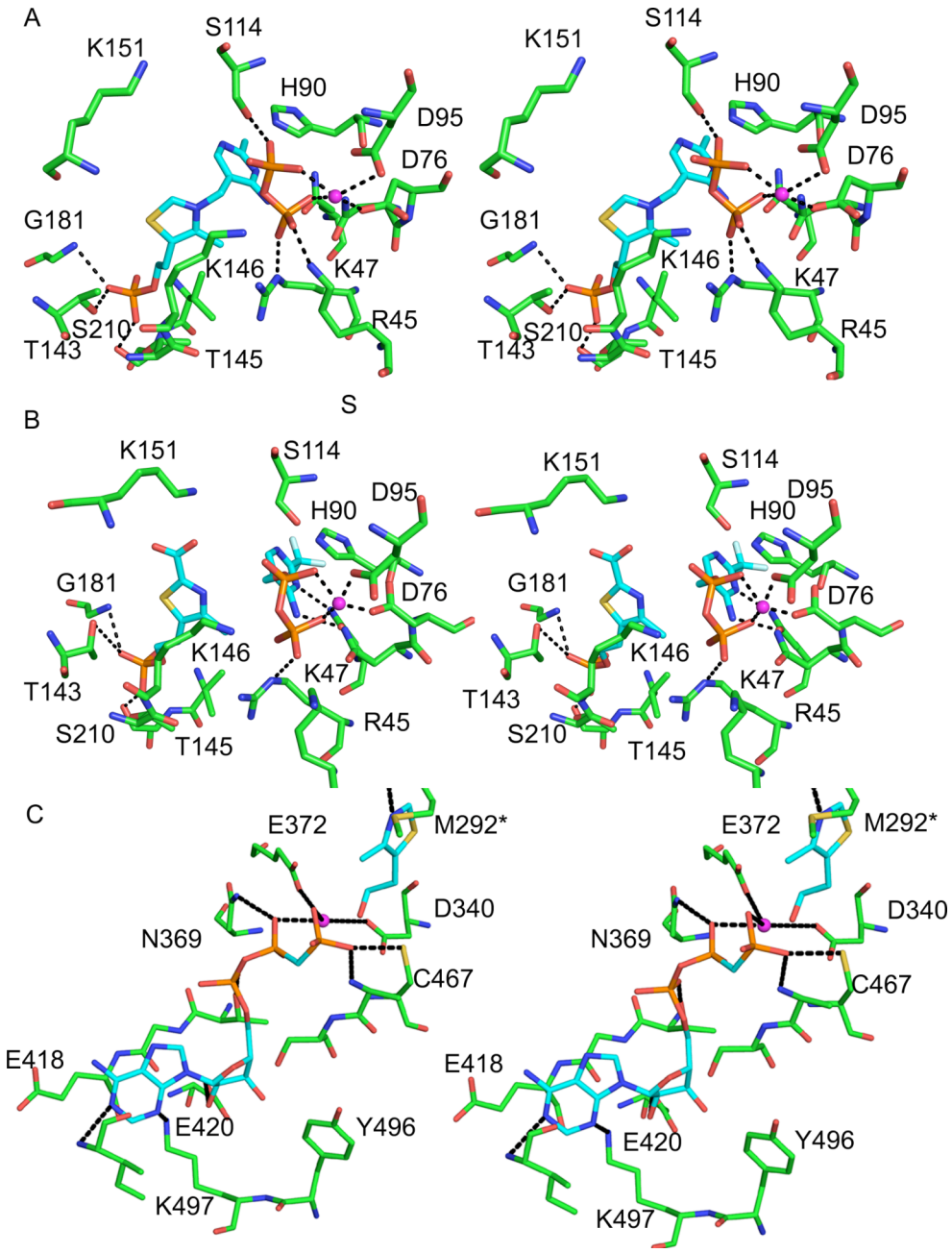
(residues 515-523) and the loop between α′9 and α′10 (residues 507-513).

Active Site of the TPS Domain. The active site of the TPS (N-terminal) domain is located in a cavity surrounded by loops 2 (residues 45-51), 4 (residues 92-97) and 6 (residues 138-154) the $(\beta\alpha)_8$ barrel. Loop 6 is the longest one and is partly disordered in the unliganded CgTHI6 structure.

The binding geometry of ThMP in the CgTHI6/ThMP complex is similar to that of the BsTPS structure (*8*). The main interactions between the protein and the ligand are shown in Figure 2.5A. The V-shaped conformation of ThMP results in part from interactions between by Ile79 and both the pyrimidine and thiazole rings. The pyrimidine moiety is buried in a mostly hydrophobic pocket formed by Tyr13, Val15, His90, Tyr134, Ile179 and Cys207 and located near the entrance of the barrel. The only hydrogen bond occurs between N3 of the pyrimidine and the side chain of Gln43. Other interactions include stacking with His90, Tyr13 and Tyr134. The thiazolium moiety binds in an extended conformation lying under loop 6. The side chain of Lys146, running parallel to the ethyl group of the thiazolium ring provides van der Waals interactions. The phosphate group forms an extensive network of hydrogen bonds with Thr143, Thr145, Gly181, Val 209 and Ser210; these residues come mostly from loop 6 and α8.

The pyrophosphate binding site in the CgTHI6/ThMP/PP complex is located near the opening of the active site and is more solvent exposed then the ThMP (Figure 2.5A). The pyrophosphate and the thiazole moiety are present on the opposite faces of the pyrimidine ring of ThMP. Interactions of ThMP are similar to those for the CgTHI6/ThMP complex. Pyrophosphate is stabilized by hydrogen bonds and electrostatic interactions. The electrostatic interactions involve, side-chains of Arg45 and Lys47 from loop 1, Lys146 from loop 6 and one Mg^{2+} ion.

Figure 2.5. The active sites of CgTHI6. (A) Stereo-diagram showing active site of the N-terminal domain of the enzyme. The product ThMP, pyrophosphate and the active site residues are shown in ball and stick presentation with carbon atoms in cyan for ligand, green for protein, nitrogen in blue, oxygen in red, sulfur yellow and phosphorus in orange. Mg^{2+} is marked by a magenta sphere. (B) Stereo-diagram showing the carboxy-thiazole binding in the active site of the N-terminal domain of the enzyme. The substrates, CO_2-Thz-P, CF_3HMP-PP, pyrophosphate and the active site residues are shown in ball and stick presentation with carbon atoms in cyan for ligand, green for protein, nitrogen in blue, oxygen in red, sulfur yellow and phosphorus in orange. Mg^{2+} is marked by a magenta sphere. (C) Stereo-diagram showing active site of the C-terminal domain of the enzyme. The substrate Thz, AMP-PCP and the active site residues are shown in ball and stick presentation with carbon atoms in cyan for ligand, green for protein, nitrogen in blue, oxygen in red, sulfur yellow and phosphorus in orange. Mg^{2+} is marked by a magenta sphere.


A
K151
S114
H90
D95
D76
G181
K146
K47
R45
S210
T143
T145
S
B
C
E372
M292*
D340
N369
C467
E418
E420
Y496
K497

The hydrogen bonding interactions occur between the pyrophosphate oxygen atoms and side-chains of Asn75, Asp76, Asp95 and Ser214. Loop 6 is partly disordered in the THI6/ThMP structure (residues 146-149 absent) and becomes ordered after pyrophosphate binding.

In the CgTHI6/ThMP/PP complex structure the Mg^{2+} is bound on the edge of the active site. The Mg^{2+} is octahedrally coordinated by the side-chains of Asp76 and Asp95 and the oxygen atoms of pyrophosphate. However in absence of pyrophosphate, the Mg^{2+} binding site is slightly shifted and the ion is tetrahedrally coordinated by side-chains of Asp76 and Asp95 and the main-chain of Ala49 and side-chain of Thr51 from the neighboring monomer.

CO_2-Thz-P is a product of the ThiK reaction and a substrate of the TPS reaction; however, in the CgTHI6/CF_3HMP-PP/CO_2-Thz-P complex (Figure 2.5B), density for CO_2-Thz-P was found only in the TPS active site. Similar interactions are observed between the thiazole moiety and the protein as for the product analogue complexes, CgTHI6/ThMP and CgTHI6/ThMP/PP. Two lysine residues are positioned near the 2-carboxy group of the thiazole ring; Lys146 and Lys151, both of them coming from loop 6. The interactions with the pyrimidine moiety and the pyrophosphate remain the same as observed in the other structures.

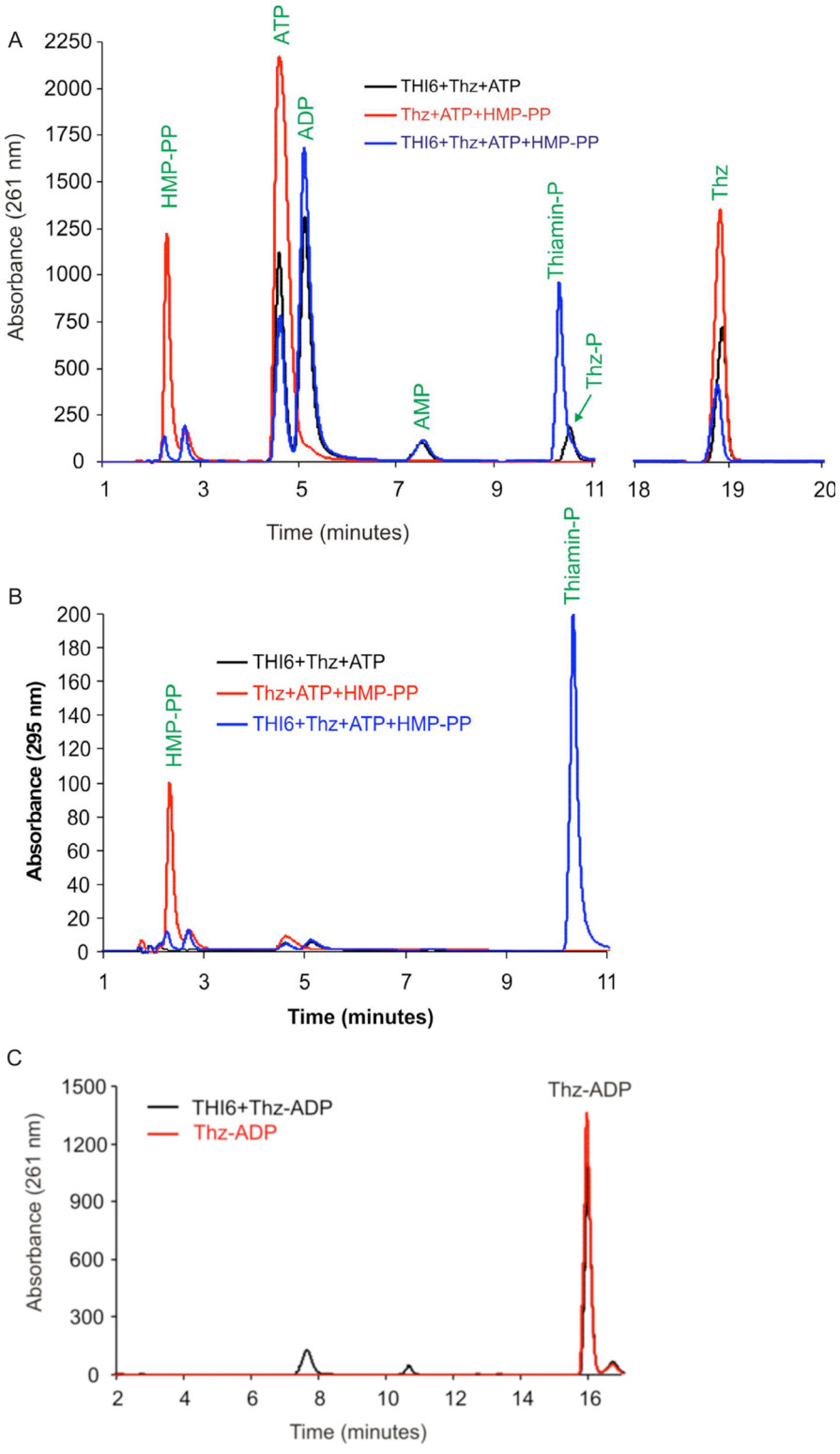
Active Site of the ThiK Binding Site. The Thz binding site in the CgTHI6/Thz/AMP-PCP is located at the interface of two domains and is rich in hydrophobic residues (Figure 2.5C). The overall active site is similar to that of BsThiK (*32*). The main interactions occur from Asn272, Val274 and Thr462 from one chain and Ala280, Leu284, Pro290 and Met292 from the other chain. The only hydrogen bond is formed between the nitrogen atom of thiazole ring and the amide nitrogen of Met292. One side of the thiazole ring is stacked against the side chain of Val274, while the other side is more solvent exposed.

The side chain of Cys467 is positioned near the hydroxyl group of thiazole.

The ATP binding site is characterized by the two complexes containing the nohydrolizable ATP analogue, AMP-PCP. The site is located near the C-terminal end of the central β-sheet and includes β′5, β′6, β′7 and β′8 (Figure 2.5C). The adenine moiety of the nucleotide is positioned under two loops, one containing a β-turn connecting β′6 and β′7 (residues 417-419), and the other containing a loop preceding α′8 (residues 451-457). The region containing residues 456-465 becomes ordered upon AMP-PCP binding. The N1 and N3 atoms of the adenine form hydrogen bonds with the Ile455 amide nitrogen and the Lys497 side chain, respectively. The ribose moiety is hydrogen bonded with the side chain of Asp420. One oxygen atom of the α-phosphate interacts with the hydroxyl group of Thr416. The side-chain of Lys367 forms a hydrogen bond with the β-phosphate, which is connected to Mg^{2+}. The γ-phosphate of the nucleotide is oriented towards the hydroxyl group of thiazole and is hydrogen bonded with the side chain of Cys467 and is coordinated with Mg^{2+}. A second Mg^{2+} interacts with the β and γ-phosphates of the nucleotide and the side-chains of the strictly conserved residues Asp340 and Glu372.

Biochemical Activities of CgTHI6. Recombinant CgTHI6 exhibited both ThiK and TPS activities (Figure 2.6A, B). When the purified protein was incubated with Thz and ATP, formation of Thz-P was noted. Incorporation of HMP-PP in the same reaction resulted in the production of ThMP. Consumption of Thz in the latter reaction was higher, presumably due to the removal of the Thz-P, product of the ThiK reaction. However, incubation of the enzyme with THI4 product ADT did not show any reaction suggesting that ADT is not a substrate for THI6 (Figure 2.6C).

Figure 2.6. Activity assays for CgTHI6. (A) HPLC analyses of the TPS and ThiK activities, absorbance measured at 261 nm. The red trace shows the control where the enzyme is not added. The black trace shows the Thz-P formation in presence of the enzyme showing the ThiK activity. The blue trace shows the ThMP formation in presence of the enzyme showing the TPS activity. (B) HPLC analyses of the TPS and ThiK activities, absorbance measured at 261 nm. The red trace shows the control where the enzyme is not added. The black trace shows the Thz-P formation in presence of the enzyme showing the ThiK activity. The blue trace shows the ThMP formation in presence of the enzyme showing the TPS activity. (C) HPLC analysis of ADT cleavage activity, absorbance measured at 261 nm. The red trace shows the control where the enzyme is not added. The black trace shows that there is no reaction in presence of the enzyme.


A
Absorbance (261 nm)
Time (minutes)
THI6+Thz+ATP
Thz+ATP+HMP-PP
THI6+Thz+ATP+HMP-PP
HMP-PP
ATP
ADP
AMP
Thiamin-P
Thz-P
Thz
B
Absorbance (295 nm)
Time (minutes)
THI6+Thz+ATP
Thz+ATP+HMP-PP
THI6+Thz+ATP+HMP-PP
HMP-PP
Thiamin-P
C
Absorbance (261 nm)
Time (minutes)
THI6+Thz-ADP
Thz-ADP
Thz-ADP

Section 2.4 Discussion

Reliability of the Medium Resolution Structure. The resolutions of the reported unliganded and complex structures range from 2.6 Å to 3.3 Å, a range considered to be medium resolution in protein crystallography. However, the current structures represent a favorable scenario. The determination of these structures involved molecular replacement using high-resolution BsThiK (1.5 Å) and BsTPS (1.25 Å) structures as starting models making the current structures were very reliable. Furthermore, the presence of six monomers in the asymmetric unit allowed the use of a six-fold NCS averaged map for model building and tight NCS restraint for refinement, which further strengthened the reliability of the structural information and resulted in a small difference between R and R-free (3-4 %).

Role of Evolution in the Formation of Bifunctional Enzyme. While the architecture of this multidomain enzyme and the chemical reactions catalyzed by each domain have been unraveled, question still remains regarding a possible cooperative interaction between the two domains. The fusion of a biosynthetic enzyme and a salvage enzyme by Nature raises the important question of the role of evolution therein. The fact that the product of the C-terminal domain, Thz-P is a substrate for the N-terminal domain explains the rationale behind the formation. In the current structural studies, no evidence is observed for the existence of a channel for the direct transfer of Thz-P from one domain to the other. However, the fusion of the two domains could be beneficial simply due to the enhancement in the local concentration of Thz-P. The conservation of the oligomeric states in the bacterial monofunctional and eukarotic bifunctional enzymes further strengthens the role of evolution in structure-function relationship.

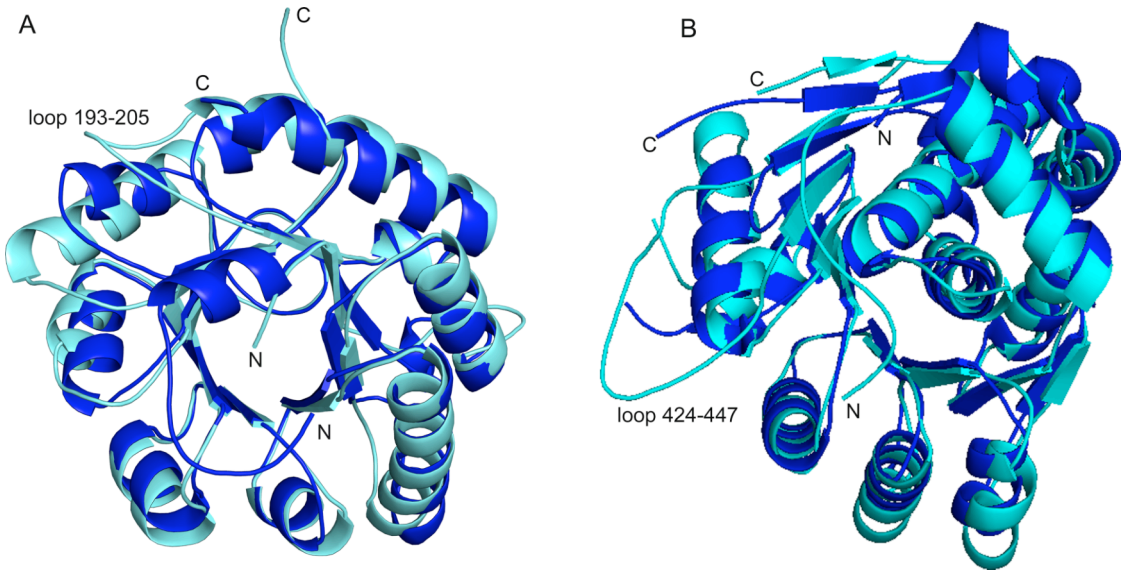


Figure 2.7. Comparison between the bifunctional CgTHI6 and its monofunctional bacterial homologues. (A) Ribbon diagram of the N-terminal domain of CgTHI6 superimposed on BsTPS monomer. CgTHI6 is colored in cyan and BsTPS in blue. (B) Ribbon diagram of the C-terminal domain of CgTHI6 superimposed on BsThiK monomer. CgTHI6 is colored in cyan and BsTPS in blue.

Structural Comparison with Bacterial TPS. TPS catalyzes the formation of ThMP from HMP-PP and Thz-P. High resolution structures of this enzyme have been reported in unliganded and complex forms from *B. subtilis* (*33*) and *P. furiosus* (PDB 1XI3). An initial structural homology search using the DALI server (*34*) indicated that BsTPS is structurally homologous to CgTHI6 having a Z score of 26.5 and RMSD 1.9 Å, based on alignment of 236 residues of the N-terminal domain of CgTHI6.

Although the N-terminal domain of CgTHI6 utilizes the same $(\beta\alpha)_8$ barrel fold as BsTPS (Figure 2.7A), a detailed structural comparison suggests that there are certain structural changes between the eukaryotic enzyme and its bacterial homologue. The significant differences occur mainly in the loop regions. The most interesting one is the presence of long loop (193-203) in the bifunctional enzyme holding the two domains together.

The active site residues of the two enzymes are mostly conserved with the same binding conformation of ThMP, PP and Mg^{2+}. A comparison of the active site residues has been listed in Table 2.4.

Table 2.4. Comparison of the active site residues of BsTPS and CgTHI6.

	BsTPS	CgTHI6
ThMP binding residues	Gln 57	Gln 43
	Thr 156	Thr 143
	Thr 158	Thr 145
	Gly 168	Gly 181
	Ser 209	Ser 210
	Ile 208	Val 209
Pyrophosphate binding residues	Ser 130	Ser 114
	Arg 59	Arg 45
	Lys 159	Lys 146
	Asn 92	Asn 75
	Lys 61	Lys 47
Mg^{2+} binding residues	Asp 93	Asp 76
	Asp 112	Asp 95

Comparison of the Dimer Interfaces of CgTHI6 with Bacterial TPS and TenI. The N-terminal domain of CgTHI6 and BsTPS both exist as dimers of $(\beta\alpha)_8$ barrels. However in the first one the barrel axes are *cis* to each other unlike in BsTPS, where, they are in *trans* orientation with respect to each other (Figure 2.8). In BsTPS dimer the major interactions occur between two symmetry related $\alpha 3$ helices running parallel to each other. Interestingly the dimeric interface of the bifunctional enzyme is similar to PfTPS, the available structure of archaebacterial TPS. The CgTHI6 dimer is also similar to BsTenI (*35*), which is an enzyme in bacterial thiamin biosynthetic pathway, showing a significant structural homology to TPS. Previous gene identification studies have suggested that *tenI* genes are paralogues of *thiE*, the gene encoding TPS (*36*). All three dimers have the barrel axes *cis* to each other (Figure 2.8). The formation of a different dimeric subunit in BsTPS is not yet well understood. One

possible reason can be considered as an artifact of the crystal packing. The occurrence of same dimeric subunit in PfTPS (archebacterium), BsTenI (eubacterium) and CgTHI6 (eukaryote) also suggests the possibility of a divergent evolution, all three enzymes originating from a common ancestor.

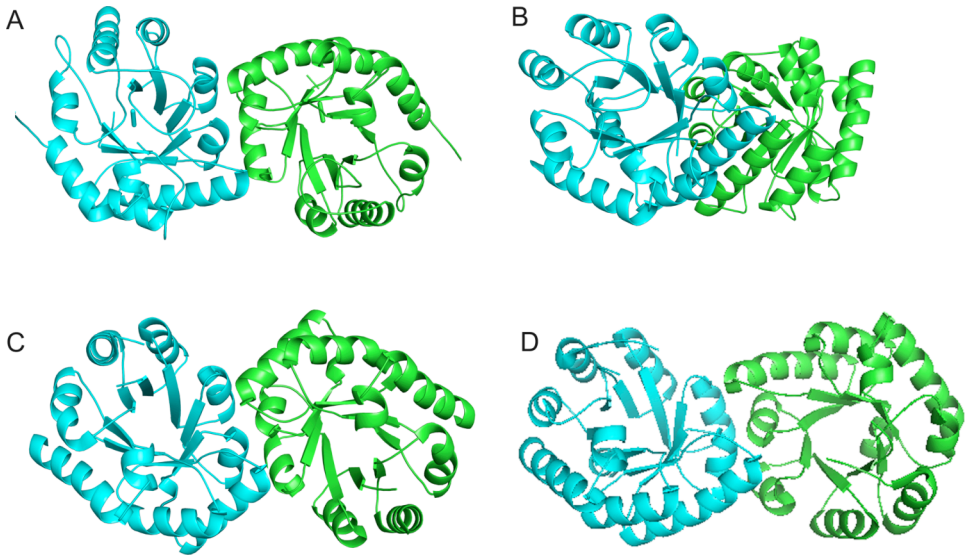


Figure 2.8. Comparison of the dimeric subunit of CgTHI6 with other enzymes. (A) Ribbon diagram showing dimer of N-terminal domain of CgTHI6. (B) Ribbon diagram showing dimer of BsTPS. (C) Ribbon diagram showing dimer of BsTenI. (D) Ribbon diagram showing dimer of PfTPS. In all figures the monomers are labeled by different colors.

Conformational Change upon Ligand Binding in TPS Domain. A comparison between the structures of the unliganded enzyme and the THI6/ThMP and THI6/ThMP/PP complexes suggests that there are some conformational changes upon ligand binding. The active site C loop composed of residues 143-153 becomes ordered upon product binding. The residues from the C loop, Thr143 and Thr145 takes part in phosphate binding of the product ThMP. The side-chain of Lys146 is involved in pyrophosphate binding. This longest loop acts as a gate near the active site, which closes upon the product binding and opens when the reaction is complete and the product is dissociated from the enzyme surface. The other active site loops A and B also show some minor changes upon product binding.

Structural Comparison with Bacterial Thiazole Kinase. ThiK catalyzes the conversion of Thz to Thz-P, in presence of ATP and Mg^{2+}. This is a salvage enzyme that provides an alternative to the regular thiazole biosynthetic pathway and enables the cell to use the recycled Thz. The crystal structures of the unliganded ThiK, ThiK/Thz and ThiK/ThiZP/ATP complexes have been reported from *B. sublitis,* in monoclinic and rhombohedral forms (*32*). Structural similarity between these two enzymes was found using DALI search, which showed a Z score of 31.6 and RMSD 1.9 Å, for 270 superimposed Cα atoms. The trimeric assembly remains conserved in the eukaryotic homologue.

The overall fold of C-terminal domain of CgTHI6 is similar to BsThiK, having a centrally located β sheet flanked by several α helices on both sides. The secondary structural elements are mostly conserved between the two enzymes (Figure 2.7B). The major structural differences are observed in the loop regions. The presence of the insertion (residues 426-446) in the bifunctional enzyme is in the form of a single turn helix, $\alpha'8_0$ (426-429) followed by an extended loop, interacting with the N-terminal domain. The loop connecting $\alpha'9$ and $\alpha'10$ is longer in bifunctional enzyme (residues

479-486) and interacts with a portion of the linker between the two domains (residues 238-241). The C-terminal helix α′12 composed of residues 256-264 in BsThiK is substituted by a single turn helix containing residues 530-533 in CgTHI6. A very interesting observation is that the loop containing residues 379-393 is disordered in our current structures. The corresponding loop containing residues 135-150 is present in the ThiK/Thz (rhombohedral crystal form) structure but disordered in ThiK/Thz-P/ATP complex structure (monoclinic crystal form) (*32*). However the structural significance of this loop is still not known.

A comparison of the active site residues between the two enzymes is shown in Table 2.5. The active sites are mostly conserved with similar binding orientations of Thz, AMP-PCP and Mg^{2+}. However, there are some minor modifications observed in the eukaryotic enzyme. The thiazole-binding loop containing residues 314-316 is slightly shifted away from the ligand in CgTHI6 compaired to the loop containing

Table 2.5. Comparison of the active site residues of BsThiK and CgTHI6.

	BsThiK	CgTHI6
Mg^{2+} Binding residues	Asp 94	Asp 340
	Glu 126	Glu 372
AMP-PCP binding residues	Arg 121	Lys 497
	Asn 123	Asn 369
	Thr 168	Thr 416
	Gly 169	Gly 417
	Glu 170	Glu 418
	Thr 191	Gly 459
	Gly 195	Ala 463
	Ala 196	Ser 464
	Tyr 225	Tyr 496
	Asp 172	Asp 420
Thiazole binding residues	Cys 198	Cys 466
	*Met 45	*Met 292

*Residues from the other chain.

residues 66-69 in BsThiK. The ATP binding site also shows some differences in the eukaryotic enzyme. A very interesting shift of ~8 Å is observed in the ATP binding loop containing residues 451-455, compared to the loop containing residues 184-188 in BsThiK. The adenine base remains more solvent exposed in the bacterial enzyme with no hydrogen bond between the adenine base and the protein. In the bifunctional enzyme the adenine ring is shielded by the loop 451-455. There are two hydrogen bonds between the ring and the protein. The N1 of the adenine base makes one hydrogen bond with the main chain nitrogen of Ile455 and the other one is between N3 of the adenine base and the side-chain of Lys497, which is replaced by Gly226 in BsThiK.

Conformational Changes Upon Thz and AMP-PCP Binding. Thz binding in the C-terminal domain induces an interesting change of space group from primitive monoclinic $P2_1$ to C-centered monoclinic $C2$, having three monomers in the asymmetric unit, composing half of the hexamer. However the whole hexamer is generated by the crystallographic 2-fold axis. The active site loop, preceeded by the single turn helix contaning residues 456-464 also becomes ordered upon ligand binding. This observation is consistent with the fact that several residues in this region make contact with the ligand.

Catalytic Mechanism. The study of the THI6/ThMP and THI6/ThMP/PP complexes show that all the active site residues are conserved between CgTHI6 and BsTPS. The binding conformations of the products, ThMP, PP and Mg^{2+} also remain same in the two active sites, indicating the involvement of a similar dissociative mechanism with the formation of pyrimidine carbocation (*33*). The binding of HMP-PP and Thz-P to the active site is followed by the closing of the C-loop, forming a stable enzyme-substrate ternary complex. The ionizaton of HMP-PP is facilitated by the stabilization of pyrophosphate as a leaving group and the delocalization of the

positive charge in the pyrimidine ring. This intermediate is trapped by the nucleophilic thiazole moiety to form thiamin phosphate, followed by the decarboxylation. The decarboxylation is facilitated by the positive charge on nitrogen of the thiazole ring.

The structural comparison of THI6/Thz-P/AMP-PCP complex and ThiK/Thz-P–/AMP-PCP complex shows that the active site residues are mostly conserved between the two enzymes with a similar kind of ligand binding. This provides an insight to the similar mechanism of action between the two enzymes. The positioning of the γ-phosphate of AMP-PCP and the Thz indicates an inline phosphate transfer. Cys467 is the probable candidate acting as a nucleophile to deprotonate the thiazole alcohol resulting in the phosphorylation. Interestingly this cystine residue is replaced by aspartate in the other known small molecule kinases, ribokinase and adenosine kinase (*37, 38*).

The substrates for bacterial TPS are CO_2-Thz-P and HMP-PP. Mechanism of the decarboxylation is not yet established in the prokaryotic system. Recent findings indicate that a carboxylated thiazole could be the substrate for thiamin phosphate synthase in the eukaryotic system as well. The active site of THI6/CF_3HMP-PP/ CO_2-Thz-P complex shows that CO_2-Thz-P binds in the TPS domain, indicating that the decarboxylation occurs in the TPS domain of the multidomain enzyme. There are two lysine residue positioned near the carboxylate group, Lys151 and Lys146, at distances 3.6 Å and 4 Å respectively. Lys146 running parallel to the ethyl group of the thiazolium ring is strictly conserved in the THI6 orthologues as well as the bacterial TPS homologues. On the other hand Lys151 does not remain conserved in the eukaryotes and is replaced by a histidine residue (His132) in BsTPS. Based on this conservation among all TPS enzymes, we suggest that Lys146 is likely to be involved in binding the negatively charged carboxylate.

Absence of ADT Cleavage Activity. The results from the activity assay clearly suggest that ADT, the THI4 product is not a substrate for THI6. This observation raises important questions about how the cleavage of the adenylyl group from ADT results in the formation of CO_2-Thz-P, a substrate for THI6. This suggests the probability of existence of another enzyme in the biosynthetic pathway.

Section 2.5. Acknowledgements

The collaborative study in this chapter focuses on the structural and biochemical aspects of the bifunctional enzyme CgTHI6. The biochemical analysis was done in Prof. Tadhg Begley's laboratory by Abhishek Chatterjee. The synthesis of the compounds, ThZ-P, HMPPP and ThZ-ADP was also done by Abhishek Chatterjee.

REFERENCES

1. Settembre, E., Begley, T. P., and Ealick, S. E. (2003) Structural biology of enzymes of the thiamin biosynthesis pathway, *Curr. Opin. Struct. Biol. 13*, 739-747.
2. Jordan, F. (2003) Current mechanistic understanding of thiamin diphosphate-dependent enzymatic reactions, *Nat. Prod. Rep. 20*, 184-201.
3. Harper, C. (2006) Thiamine (vitamin B1) deficiency and associated brain damage is still common throughout the world and prevention is simple and safe!, *Eur. J. Neurol. 13*, 1078-1082.
4. Begley, T. P., Downs, D. M., Ealick, S. E., McLafferty, F. W., Van Loon, A. P., Taylor, S., Campobasso, N., Chiu, H. J., Kinsland, C., Reddick, J. J., and Xi, J. (1999) Thiamin biosynthesis in prokaryotes, *Arch. Microbiol. 171*, 293-300.
5. Kriek, M., Martins, F., Leonardi, R., Fairhurst, S. A., Lowe, D. J., and Roach, P. L. (2007) Thiazole synthase from *Escherichia coli*: an investigation of the substrates and purified proteins required for activity *in vitro*, *J. Biol. Chem. 282*, 17413-17423.
6. Chatterjee, A., Li, Y., Zhang, Y., Grove, T. L., Lee, M., Krebs, C., Booker, S. J., Begley, T. P., and Ealick, S. E. (2008) Reconstitution of ThiC in thiamine pyrimidine biosynthesis expands the radical SAM superfamily, *Nat. Chem. Biol. 4*, 758-765.
7. Cheng, G., Bennett, E. M., Begley, T. P., and Ealick, S. E. (2002) Crystal structure of 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase from *Salmonella typhimurium* at 2.3 Å resolution, *Structure 10*, 225-235.
8. Chiu, H. J., Reddick, J. J., Begley, T. P., and Ealick, S. E. (1999) Crystal

structure of thiamin phosphate synthase from *Bacillus subtilis* at 1.25 Å resolution, *Biochemistry 38*, 6460-6470.

9. McCulloch, K. M., Kinsland, C., Begley, T. P., and Ealick, S. E. (2008) Structural studies of thiamin monophosphate kinase in complex with substrates and products, *Biochemistry 47*, 3810-3821.
10. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. (2002) Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms, *J. Biol. Chem. 277*, 48949-48959.
11. Jurgenson, C. T., Chatterjee, A., Begley, T. P., and Ealick, S. E. (2006) Structural insights into the function of the thiamin biosynthetic enzyme Thi4 from *Saccharomyces cerevisiae*, *Biochemistry 45*, 11061-11070.
12. Chatterjee, A., Jurgenson, C. T., Schroeder, F. C., Ealick, S. E., and Begley, T. P. (2006) Thiamin biosynthesis in eukaryotes: characterization of the enzyme-bound product of thiazole synthase from *Saccharomyces cerevisiae* and its implications in thiazole biosynthesis, *J. Am. Chem. Soc. 128*, 7158-7159.
13. Zeidler, J., Sayer, B. G., and Spenser, I. D. (2003) Biosynthesis of vitamin B1 in yeast. Derivation of the pyrimidine unit from pyridoxine and histidine. Intermediacy of urocanic acid, *J. Am. Chem. Soc. 125*, 13094-13105.
14. Kawasaki, Y. (1993) Copurification of hydroxyethylthiazole kinase and thiamine-phosphate pyrophosphorylase of *Saccharomyces cerevisiae*: characterization of hydroxyethylthiazole kinase as a bifunctional enzyme in the thiamine biosynthetic pathway, *J. Bacteriol. 175*, 5153-5158.
15. Nosaka, K., Nishimura, H., Kawasaki, Y., Tsujihara, T., and Iwashima, A. (1994) Isolation and characterization of the THI6 gene encoding a bifunctional thiamin-phosphate pyrophosphorylase/hydroxyethylthiazole kinase from *Saccharomyces cerevisiae*, *J. Biol. Chem. 269*, 30510-30516.

16. Lawhorn, B. G., Mehl, R. A., and Begley, T. P. (2004) Biosynthesis of the thiamin pyrimidine: the reconstitution of a remarkable rearrangement reaction, *Org. Biomol. Chem. 2*, 2538-2546.

17. Zamyatina, A., Gronow, S., Oertelt, C., Puchberger, M., Brade, H., and Kosma, P. (2000) Efficient Chemical Synthesis of the Two Anomers of ADP-L-glycero- and D-glycero-D-manno-Heptopyranose Allows the Determination of the Substrate Specificities of Bacterial Heptosyltransferases This work was supported by grants from FWF (grant nos. P11449-MOB and P13843-CHE), *Angew Chem Int Ed Engl 39*, 4150-4153.

18. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem. 72*, 248-254.

19. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.

20. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool, *J. Mol. Biol. 215*, 403-410.

21. Matthews, B. W. (1968) Solvent content of protein crystals, *J. Mol. Biol. 33*, 491-497.

22. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Improved sensitivity of profile searches through the use of sequence weights and gap excision, *Comput. Appl. Biosci. 10*, 19-29.

23. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D 54*, 905-921.

24. Collaborative Computational Project-Number 4. (1994) The CCP-4 suite: programs for protein crystallography, *Acta. Crystallogr. D 50*, 760-763.

25. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.

26. Kleywegt, G. J., and Jones, T. A. (1996) xdlMAPMAN and xdlDATAMAN-programs for reformatting, analysis, and manipulation of biomacromolecular electron-density maps and reflection datasets, *Acta Crystallogr. D 52*, 826-828.

27. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination, *Acta Crystallogr. D 58*, 1948-1954.

28. van Aalten, D. M., Bywater, R., Findlay, J. B., Hendlich, M., Hooft, R. W., and Vriend, G. (1996) PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules, *J Comput Aided Mol Des 10*, 255-262.

29. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript, *Bioinformatics 15*, 305-308.

30. DeLano, W. L. (2002), The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA.

31. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol. 372*, 774-797.

32. Campobasso, N., Mathews, II, Begley, T. P., and Ealick, S. E. (2000) Crystal structure of 4-methyl-5-β-hydroxyethylthiazole kinase from *Bacillus subtilis* at 1.5 Å resolution, *Biochemistry 39*, 7868-7877.

33. Peapus, D. H., Chiu, H. J., Campobasso, N., Reddick, J. J., Begley, T. P., and Ealick, S. E. (2001) Structural characterization of the enzyme-substrate,

enzyme-intermediate, and enzyme-product complexes of thiamin phosphate synthase, *Biochemistry 40*, 10103-10114.

34. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol. 233*, 123-138.

35. Toms, A. V., Haas, A. L., Park, J. H., Begley, T. P., and Ealick, S. E. (2005) Structural characterization of the regulatory proteins TenA and TenI from *Bacillus subtilis* and identification of TenA as a thiaminase II, *Biochemistry 44*, 2319-2329.

36. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. (2002) Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms, *J. Biol. Chem 277*, 48949-48959.

37. Mathews, II, Erion, M. D., and Ealick, S. E. (1998) Structure of human adenosine kinase at 1.5 Å resolution, *Biochemistry 37*, 15607-15620.

38. Sigrell, J. A., Cameron, A. D., Jones, T. A., and Mowbray, S. L. (1998) Structure of *Escherichia coli* ribokinase in complex with ribose and dinucleotide determined to 1.8 Å resolution: insights into a new family of kinase structures, *Structure 6*, 183-193.

CHAPTER 3

STRUCTURAL STUDIES ON *ZEA MAIZE* THIAZOLE SYNTHASE: ROLE OF VAL211

Section 3.1 Introduction

Thiamin (vitamin B_1) consisting of a thiazole moiety linked to a pyrimidine ring, is an essential molecule in all forms of life. The active form, thiamin pyrophosphate (TPP) plays an important role as a cofactor for several enzymes involved in carbohydrate and amino acid metabolism (*1*). Some important new functions of the phosphate esters of thiamin have been recently suggested. The binding of thiamin monophosphate (ThMP) to mRNA domains called "riboswitches" regulate the expression of thiamin biosynthetic genes at the transcriptional level (*2*). Recent findings have shown that ThMP also plays a major role in repairing DNA damage and increasing mitochondrial stability (*3, 4*). Thiamin triphosphate has been suggested to play a role in physiology of the nervous system (*5*).

Thiamin is biosynthesized by prokaryotes and some eukaryotes, including fungi and plants. Vertebrates including human, need to obtain this essential cofactor through diet and deficiency can lead to a disease called beriberi characterized by disturbances of the central nervous system (*6*). In plants, several mutants defective in different stages of thiamin biosynthesis have been reported. Recent studies on *Zea maize* showed that a mutation in the thiazole biosynthetic gene *thi1_2* results a phenotype called Bladekiller1 (*blk1*) exhibiting premature termination of influorescence and axillary meristems and poor development of leaf blades (John Woodward, private communication). The point mutation occurs in a highly conserved valine residue and replaces it with methionine. This is caused by a G to A nucleotide mutation, common in plants that have been treated with ethylmethylsulfonate. Two isoforms of *Zea maize thi1* (*thi1_1* and *thi1_2*) have been functionally assigned as

thiazole synthase based on their high sequence similarity with previously characterized THI4 from *Saccaromyces cerevisiae* (ScTHI4) (*7*). The isoforms are 90% identical with tight conservation of the putative active site residues.

Thiamin biosynthesis has been studied extensively in prokaryotic systems both structurally and mechanistically. 4-methyl-5-(hydroxyethyl)thiazole phosphate (Thz-P) and 4-amino-5-(hydroxymethyl)-2-methylpyrimidine pyrophosphate (HMP-PP) are synthesized through separate routes and then coupled together to form ThMP. A final phosphorylation of ThMP leads to the formation of the biologically active cofactor. The prokaryotic thiazole biosynthetic pathway involves a complex oxidative condensation (Figure 3.1). In the first step, glyceraldehyde-3-phosphate and pyruvate are ligated by the enzyme 1-deoxy-D-xylulose-5-phosphate synthase (DXS) to form 1-deoxy-D-xylulose-5-phosphate (DXP) (*8*). In the following step, the sulfur carrier protein ThiS is adenylated by ThiF, followed by a sulfur transfer by ThiI and IscS (*E. coli*) or NifS (*B. subtilis*) to form a thiocarboxylated C-terminus (*9*). Glycine is oxidized in *B. subtilis* by ThiO (*10*) or tyrosine in *E. coli* by ThiH (*11*) to give dehydroglycine. Dehydroglycine is then coupled with DXP and the thiocarboxylated ThiS, by thiazole synthase ThiG to form thiazole phosphate carboxylate tautomer (*12*). The thiazole tautomer is then aromatized by TenI (*B. subtilis*) to form the thiazole phosphate carboxylate (CO_2-Thz-P) (*13*).

The biosynthesis of thiamin in eukaryotes involves a completely different biochemical pathway and has only beginning to emerge. A single enzyme, known as THI4 (in fungi) and THI1 (in plants) catalyzes the thiazole biosynthetic pathway in eukaryotes (*14*). Nicotinamide adenine dinucleotide (NAD), glycine and a yet unknown sulfur source are assembled together to form the final product adenosine diphospho-5-(β-ethyl)-4-methyl-thiazole-2-carboxylic acid (ADT) in a complicated

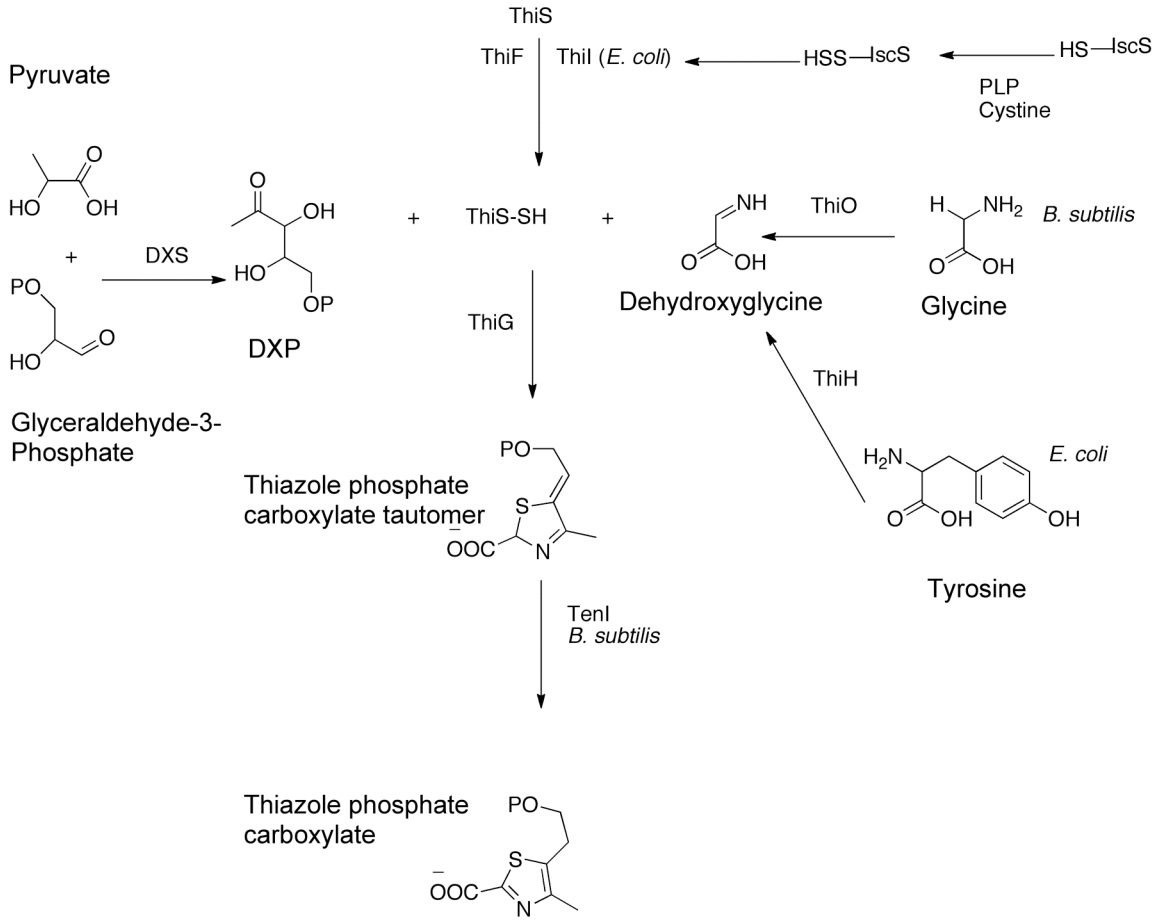


Figure 3.1. Thiazole biosynthesis in prokaryotes.

biochemical pathway (*15*). The trapping and characterization of three adenylated metabolites (A, B and C in Figure 4.2) bound to the enzyme, at different stages of the pathway, have provided a mechanistic profile for thiazole formation in yeast (*15*).

THI4 is dual functional enzyme also involved in mitochondrial DNA repair (*3*). The previously determined crystal structures of ScTHI4 and *Arabidopsis thaliana* THI1 (AtTHI1) show that this enzyme exists as a homooctamer (*16, 17*). In both the structures the product ADT is bound in the enzyme active site. The structure of ScTHI4 shows a *cis*-proline residue (121), hydrogen bonded to the phosphate group of the ADT moiety, suggesting that isomerization of the proline might lead to product release.

The enzyme from *Neurospora crassa* is highly expressed (1.5% of the entire proteome) and interacts with a *cis-trans* prolyl isomerase, further supporting this hypothesis (*18*).

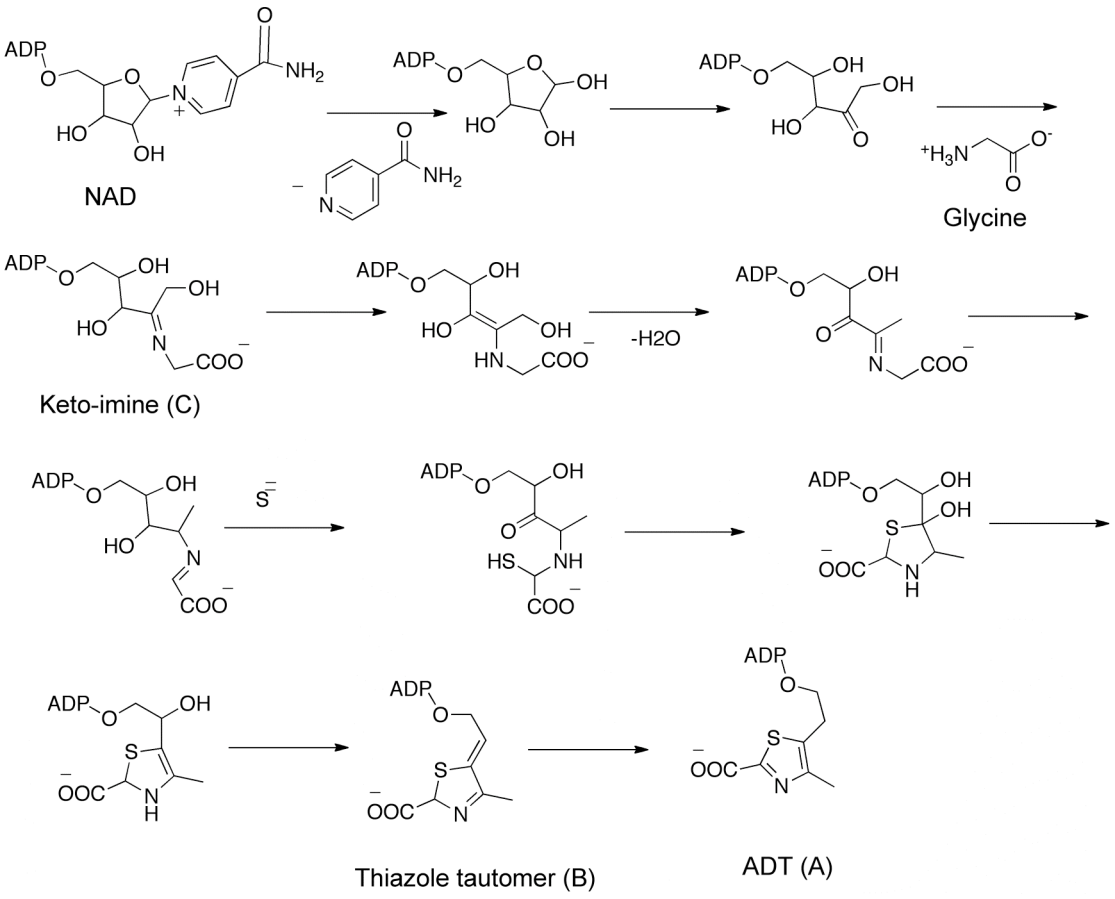


Figure 3.2. Thiazole biosynthesis in eukaryotes.

Here we report the crystal structure of wild type *Zea maize* THI1_2 (ZmTHI1) with the product ADT bound in the active site at 2.5 Å. Biochemical studies show that the wild-type enzyme purifies with the adenylated metabolites (A, B and C, Scheme 2) but the mutant does not. An assay demonstrating the reconstitution of activity has been performed with the wild type enzyme in presence of NAD, glycine and Fe^{2+}.

However, such reconstitution failed in case of the mutant. Dynamic light scattering (DLS) has been used to characterize the octameric state of the wild type enzyme in the solution state. Furthermore different mutants have been designed and studied in order to investigate the role of Val211 in structural and functional aspects of ZmTHI1.

Section 3.2 Experimental

Chemical Reagents. M9 minimal medium, L-amino acids, magnesium sulfate, calcium chloride, Kanamycin, calcium acetate, isopropyl alcohol and the buffer components were all purchased from Sigma-Aldrich. Ni-NTA resin was obtained from Qiagen (Valencia, CA). HPLC grade solvents were obtained from Fisher Scientific.

Genetic Mapping of blk1. An F2 mapping population was generated by crossing thi1_2-*blk1* heterozygotes introgressed into B73 nine times onto the MO17 inbred. Known SSR and INDEL markers for 3L were available at maizegdb.org and used to delimit the mapping interval. Sequence available at maizesequence.org and magi.plantgenomics.iastate.edu/ were used identify candidate genes and build primers for amplifying genomic regions. Sequence polymorphisms between MO17 and the *thi1_2-blk1* progenitor were used to design CAPS markers that further narrowed the mapping interval. *thi1_2* (GRMZM2G074097) was amplified by PCR using primers thi2-cDNA-f1 and thi2-cDNA-R1, which flank the predicted coding region. This work was done by John Woodward in Prof. Michael Scanlon's laboratory.

Cloning of ZmTHI1 and its Mutants. The ZmTHI1 clones were prepared by Cornell Protein Facility, using standard method of DNA manipulation (*19, 20*). Plasmid DNA was purified using the Qiagen Miniprep kit. *E. scoli* strain MachI (Invitrogen) was used for transformations during plasmid construction, propagation and storage. The first 47 N-terminal residues were chopped off because they encoded signaling sequence. Site-directed mutagenesis was performed by a standard PCR

protocol uring Pfu Turbo DNA polymerase per the manufacturer'e instructions (Invitrogen). The primers used for all constructs are listed in Supplemental Table 1.

Expression and Purification of ZmTHI1 and its Mutants. The wild-type and mutant ZmTHI1 genes were cloned into a modified pET28 overexpression vector encoding an N-terminal hexahistidine tag with a TEV protease cleavage site. The plasmid was transformed into *Escherichia coli* (Novagen) cell line BL21(DE3). A single colony was used to inoculate a 10 mL overnight culture, grown in Luria-Bertani media at 37 ºC supplemented with 60 μg/mL kanamycin. The saturated starter culture was then added to 1L of LB media, containing the same antibiotics and grown at 37 ºC with shaking at 200 rpm until the OD_{600} reached 0.7. At this point the culture was cooled to 15 ºC for 45 mins and induced with 0.1 mM isopropyl-1-β -D thiogalactopyranoside (IPTG) for 16 – 18 h. Cells were harvested by centrifugation at 9000 *g* for 15 min at 4 ºC using an Avanti J-251 centrifuge (Beckman) and frozen at -20 ºC for later use.

All purification steps were carried out at 4 °C. The frozen cell pellet from 2L ofculture was thawed and resuspended in 50 mL of lysis buffer [50 mM Tris (pH 8.0), 300 mM NaCl, 10 mM imidazole] with the addition of Protease Inhibitors (Roche) and lysozyme (1 mg/mL). The cells were lysed by sonication in an ice-water bath. A Sonic Dismembrator 500 from Fischer Scientific was used with a 5 s pulse followed by 10 s of cooling over a period of 1 minute and repeated twice . Cellular debris was removed by centrifugation at 35000 *g* for 50 min. The clarified lysate was then loaded onto a 2 mL Ni-NTA column (Qiagen) pre-equilibriated with 50 mL of the lysis buffer. The column was then washed with 100 mL of wash buffer [50 mM Tris (pH 8.0), 300 mM NaCl, 60 mM imidazole] to remove any nonspecific binding contaminants. ZmTHI1 was eluted with elution buffer [50 mM Tris (pH 8.0), 300 mM NaCl, 250 mM imidazole]. The elution was monitored by measuring protein

concentration via Coomassie Plus Protein Assay Reagent (Pierce). The proteins were further purified by gel filtration chromatography. A Hi-Load Superdex 200 column (26/60, GE Healthsciences) was equilibrated with 50 mM Tris buffer (pH 8.0) and 100 mM NaCl. Pure protein was eluted at a flow rate of 1 mL/min while monitoring the $A_{280\ nm}$ and collecting 2 mL fractions. The fractions containing the protein as evidenced by SDS-PAGE gel analysis were combined and concentrated to 10 mg/mL using an Amicon Ultra centrifugal device with a 10 kDa MWCO (Millipore). The purity of the proteins was found to be greater than 95 % assessed by Coomassie-stained SDS-PAGE analysis (data not shown).

Crystallization of ZmTHI1. The crystallization experiments were conducted at 22 ºC using hanging-drop vapor diffusion by combining 1 μL of wild type ZmTHI1 solution and 1 μL of reservoir solution. Initial crystallization conditions were determined using sparse matrix screens, Crystal Screen 1 and 2 (Hampton research). The final optimized condition contained 0.1 M MES (pH 6-6.3), 0.2-0.25 M calcium acetate and 10%-12% isopropyl alcohol. Diamond shaped crystals, 200-300 μm long, appeared in 6-8 days. 10% glycerol added to the mother liquor was used as a cryoprotectant.

The V211A mutant was crystallized under similar condition (0.1 M MES, pH 6.1, 0.2 M calcium acetate, 12% isopropyl alcohol). The crystals grew to the size of 400 μm in 3-5 days. 12% glycerol was used as cryprotectant.

Data Collection and Processing. All data were collected at the NE-CAT beamlines 24-ID-E or 24 -ID-C at the Advanced Photon Source at Argonne National Laboratory using a Quantum315 X-ray detector (Area Detector Systems Corporation). A beam with 1 sec exposure time, 1º oscillation and 400 mm of crystal to detector distance were used to record the diffraction images. The data were indexed, integrated and scaled using HKL2000 program suite (*21*) (Table 3.1).

Table 3.1. Data collection statistics. Values for the highest resolution shell are given in parentheses.

	Wild type ZmTHI1	V211A ZmTHI1
Wavelength (Å)	0.9795	0.9795
Resolution (Å)	2.5	2.04
Space group	*I*4	*I*4
Unit Cell		
a (Å)	95.9	96.0
c (Å)	130.9	131.2
Matthews coefficient	2.3	2.3
% solvent	47	47
mol/asym	2	2
Reflections	69378	158059
Unique reflections	19530	35797
Completeness (%)	98.8 (90.5)	95.2 (98.5)
Rsym *(%)	12.5 (31.4)	7.5 (36.5)
I/σ	11.4 (2.9)	18.4 (5.0)
Redundancy	3.1 (2.8)	4.4 (4.1)

*$R_{sym} = \Sigma\Sigma_i \left| I_i - \langle I \rangle \right| / \Sigma \langle I \rangle$, where <I> is the mean intensity of the N reflections with intensities I_i and common indices *h,k,l*.

Structure Determination and Refinement. The Matthews number (*22*) suggested that the asymmetric unit contains two monomers with solvent content of 45%. BLAST searches (*23*) showed that ZmTHI1 is about 85% identical to AtTHI1 (PDB 1RP0). Molecular replacement was carried out with the dimer of AtTHI1 as a search model with all water and ligand atoms removed. The Crystallography and NMR systems (CNS) program suite (*24*) was used to calculate the rotation and translation functions with a 4 Å resolution cutoff. All model building was manually done using the computer program COOT (*25*). The rounds of refinements were performed using REFMAC (*22*) and involved successive rounds of TLS and restrained refinement and model rebuilding. Tight non-crystallographic symmetry restraints (NCS) were used during the initial rounds and were gradually relaxed during the final stages of the refinement. Two-fold NCS averaged electron density maps and averaged composite omit maps were used for all stages of model building. The ligand parameter and topology files were generated using PRODRG server (*26*).

The V211A dataset showed ~35% merohedral twinning predicted by the program Xtriage in PHENIX suite (*27*). As a result, refinements were performed using REFMAC under the correct twin operator and few rounds gave reasonable R/R_{free}. A summary of the refinement statistics is shown in Table 3.2.

Table 3.2. Refinement statistics.

	Wild type ZmTHI1	V211A ZmTHI1
Resolution (Å)	2.5	2.04
No of protein atoms	4160	3984
No of water atoms	441	234
No of ligand atoms	76	-
RMSD		
Bonds (Å)	0.008	0.007
Angles (deg)	1.156	1.144
R factor [a] (%)	21.8	16.3
R_{free} [b] (%)	26.4	21.7
Ramachandran plot		94.1
Most favored (%)	93.9	5.6
Additional allowed (%)	5.8	0.3
Generously allowed (%)	0.3	0.0
Disallowed (%)	0.0	0.0

[a] R factor = $\sum_{hkl} \left| \left| F_{obs} \right| - k \left| F_{cal} \right| \right| / \sum_{hkl} \left| F_{obs} \right|$ where F_{obs} and F_{cal} are observed and calculated structure factors, respectively. [b] For R_{free} the sum is extended over a subset of reflections (5%) excluded from all stages of refinement.

HPLC Analysis of ZmTHI1 and its Mutants for Associated Metabolites. HPLC analyses of the bound metabolites were performed using a Supelco LC-18 (150x4.6 mm, 3 μm ID) reverse phase column. HPLC analyses were carried out in an Agilent 1200 instrument equipped with an inline diode-array detector. The enzyme was denatured by heat (100 °C, 60 s) and the reaction mixtures were rapidly cooled on ice. The precipitated protein was removed by centrifugation and the supernatant was filtered through a 10 kDa MWCO filter (Microcon YM-10, Millipore). 100 μL of the filtrate was analyzed by reversed phase HPLC. The following linear gradient was used at 1 mL/min flow rate (solvent A is water, solvent B is 100 mM KP_i, pH 6.6, solvent C is methanol): 0 min 100% B, 3 min 10% A 90% B, 10 min 25% A 60% B 15% C, 14 min 25% A 60% B 15% C, 19 min 30% A 40% B 30% C, 21 min 100% B, 30 min 100% B. The absorbance at 261 nm was monitored. Different compounds in the chromatogram were identified by co-migration with ScTHI4 associated metabolites. This work was performed by Dinuka Abeydeera in Prof. Tadhg Begley's Laboratory.

Expression and Purification of Metabolite Free ZmTHI1. The wild type enzyme purified with bound metabolites and was in an inactive form. In order to obtain the active enzyme the metabolite free ZmTHI1 was expressed and purified. The cell-growing minimal media was prepared by autoclaving 22.5 g of M9 minimal salt in 2 L solution in H_2O. 25 mL of 50 % (w/v) sterilized glucose, 4 mL of 1 M sterilized $MgSO_4$, 0.2 mL of 1 M sterilized $CaCl_2$, and 60 mg/L Kanamycin were added to the media upon cooling. A 20 mL overnight starter culture was used to inoculate 1 L M9 minimal media prepared as described above. The cells were grown at 37 °C, with shaking to an OD_{600} of 0.6, at which point they were induced with 1 mM IPTG. The temperature was reduced to 15 °C and the cells were additionally grown for 16-18 h. The cultures were then harvested by centrifugation at 9000 *g* for

15 min at 4 ºC using an Avanti J-251 centrifuge (Beckman) and the pellets were frozen at -20 ºC for future use.

To get the metabolite free protein, three round of gel filtrations were performed, using a Hi-Load Superdex 200 column (26/60, Amersham), equilibrated with 50 mM Tris buffer (pH 8.0) and 100 mM NaCl. The fractions containing the protein as evidenced by SDS gel analysis were combined and concentrated to 10 mg/mL using an Amicon Ultra centrifugal device with a 10 kDa MWCO (Millipore). The protein preparation was then used to set up activity assays. This work was performed by Dinuka Abeydeera in Prof. Tadhg Begley's Laboratory.

In vitro Reconstitution of Wild Type ZmTHI1 Activity. To the wild type protein (230 μM) expressed in M9 minimal media was incubated with NAD (0.5 mM final concentration), glycine (0.5 mM final concentration), and freshly prepared $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ (0.5 mM final concentration), in the glove box (anaerobic atmosphere) at room temperature. After 1 hr, an aliquot of 100 μL from the reaction mixture was quenched by adding into 100 μL of 8 M guanidine hydrochloride. The mixture was filtered through a 10 kDa MWCO filter (Microcon YM-10, Millipore). An injection volume of 50 uL was analyzed by the HPLC as previously described.

Dynamic Light Scattering Studies on ZmTHI1. DLS studies were performed on wild type and mutant ZmTHI1 using the instrument Zetasizer Nano-ZS (Malvern). The underlying principle involves measurement of the intensity of the scattered light and relating that to the size of the particle. The samples were illuminated with a red laser beam (660 nm) and the fluctuation of the scattering intensity was measured using an optical angle of 173º. The Zetasizer software was used to calculate the size distribution based on intensity. Samples were prepared in 10 mM Tris (pH 8.0), 100 mM NaCl and 4 mM DTT (buffer 1), or 10 mM Tris (pH 8.0), 100 mM NaCl and 2 % β-octyl-glucoside (buffer 2). Three different concentrations: 0.5, 1.0 and 1.5 mg/mL

were used for each sample. The average diameter was then used to calculate the approximate molecular mass.

Figure Preparation. All figures were prepared using Chemdraw and PYMOL (*28*).

Section 3.3 Results

Mapping of blk1 on Chromosome 3L. The *blk1* mutation was mapped to contig 149 of the maize agarose FPC map by linkage to SSR marker umc2152 and the indel marker IDP351. A population of 686 mutants was used to fine map the mutation to a 0.23 cM interval delimited by CAPS marker CAPSK03 (2 recombinants) and CAPSN05 (1 recombinant). This region constitutes approximately 300kb from three contiguous sequenced BACs. Analysis of the sequenced region revealed seven candidate ORFs. Each ORF was evaluated for sequence polymorphisms that would disrupt the predicted gene product. A G to A point mutation that results in a predicted missense mutation from Val211 to Met211 in the encoded protein was identified in the thiamine biosynthetic gene (*thi1_2*) (GRMZM2G074097) indicating that *blk1* is a mutant allele of *thi1_2*.

Overall Structure of ZmTHI1. The crystal structure of ZmTHI1 shows that the enzyme exists as an octamer composed of eight identical subunits (Figure 3.3a and b). The asymmetric unit contains two molecules and the homooctamer, generated through crystallographic symmetry. The octameric assembly can also be viewed as a tetramer of dimers having a shape of bi-layered ring with flattened sides. The inner and outer diameters of the ring are 30 Å and 100 Å respectively, with a thickness of 70 Å. The dimeric interface has a buried surface area of 3485 $Å^2$ with primarily hydrophobic interactions (*29*). The main dimeric interface is formed by two long antiparallel helices (α4), coming from each monomer composed of residues 166-177 (Figure 3.3c). The only hydrogen bond between the adjacent monomers is found between

Ser175 and Thr172. The second dimeric interface includes a portion of α1 from one monomer, composed of residues 72-77, interacting with two short antiparallel strands β3 (residues 155-158) and β4 (residues 160-164), mainly consisting of hydrophobic and Van der Waals interactions. The interfaces between the dimers have a buried surface area of 1000 $Å^2$ and the interactions occur in mainly two regions involving two molecules of the dimer. The residues 71-76 from α1 of chain A of one dimer (AB), interacts with residues 79-86, the remaining portion of α1 from chain C of the other dimer (CD). The second interface includes the loop joining the strands β2 and β3, consisting of residues 131-134 from chain A interacting with β4 containing residues 220-224 from chain D the other dimeric unit.

Monomeric Structure of ZmTHI1. The monomeric core of ZmTHI1 shows a αβ fold where the centrally located six stranded mixed β sheet has the topology: β5↑β2↑β1↑β9↑β11↑β10↓ (Figure 3.3d). The central sheet is flanked by a bundle of helices containing α4, α2, α9 and α7 on one side and an antiparallel sheet containing β6↑β7↓ β8↑ on the other side. The monomeric chain contains one 3_{10} helix, connecting strands β7 and β8 and composed of residues 164-167. The longest helix α1, consisting of residues 70-87 is found near the surface of the octameric structure. Seven residues in the N-terminus are found to be disordered along with twenty residues at the C-terminal end and thus not included in the model.

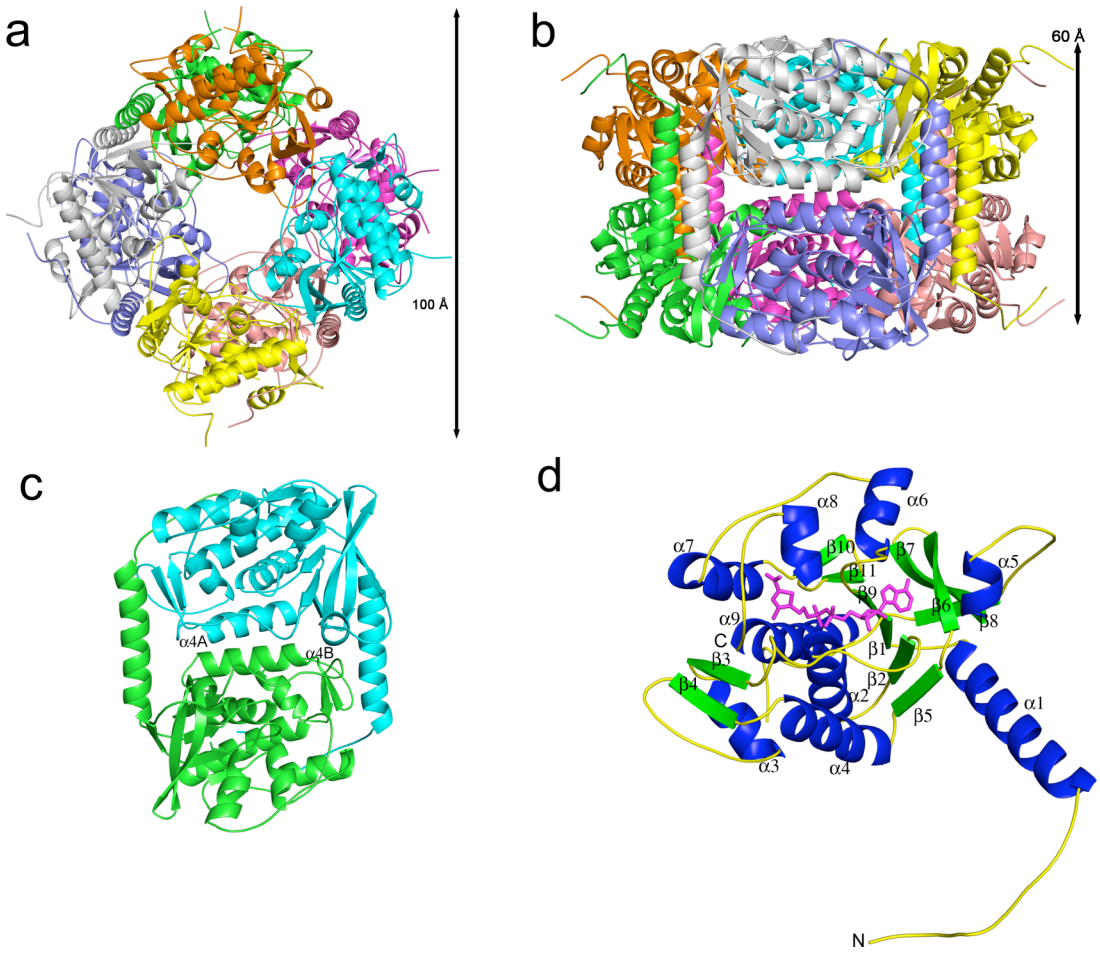


Figure 3.3. Structure of ZmTHI1. (A) Cartoon representation of the quaternary structure of ZmTHI1 octamer viewed through central cavity. (B) Cartoon representation of the quaternary structure of ZmTHI1 octamer rotated 90 ° about the horizontal axis relative to the orientation in panel A. Both the structures are color coded by monomers. (C) Cartoon representation of the dimeric subunit of ZmTHI1 color coded by monomer. The α4 helices are labeled in each monomer showing the main dimer interface. (d) Cartoon representation of ZmTHI1 monomer, colored by secondary structural elements. The α helices are colored in blue, β strands in green and loops in yellow. The ligand is shown in ball and stick representation and colored in magenta.

Active Site of ZmTHI1. The crystal structure of the product complex maps out the active site of ZmTHI1 (Figure 3.4). The active site is located near the inner side of the ring structure formed by the octameric assembly. The protein contains a signature Gly-Ala-Gly-Ser-Ala-Gly motif (residues 93-98) characteristic of a dinucleotide binding domain. The adenine ring is stabilized by a herringbone interaction with Phe241. The amide nitrogen atom and carbonyl oxygen atom of Val191 form hydrogen bonds with N1 and N6 of adenine. The 2′ and 3 ′ hydroxyl groups of the ribose moiety are anchored in the active site by forming hydrogen bonds with the carboxylate group of Glu118. The O1 from the α phosphate group is hydrogen bonded to the amide nitrogen atom of Gly126. The O1 of the β phosphate group forms a hydrogen bond with the amide nitrogen of Ala97.

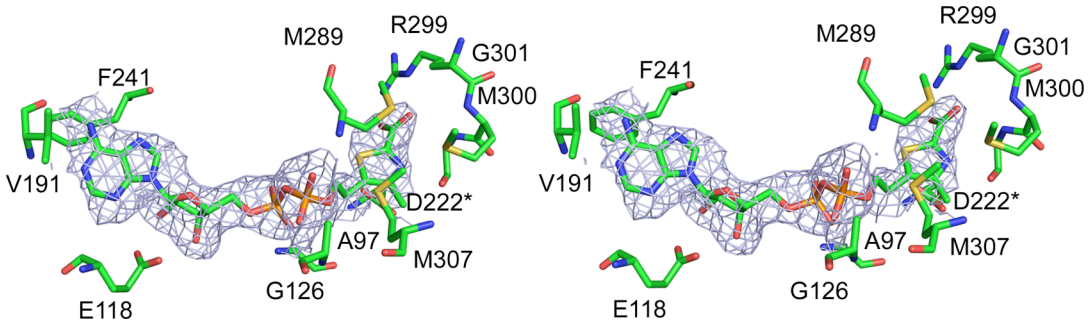


Figure 3.4. Stereo-diagram showing the active site of ZmTHI1. The product ADT and the active site residues are shown in ball and stick representation. The carbon atoms are colored in green, nitrogen in blue, oxygen in red, sulfur yellow and phosphorus in orange. All residues come from the same monomer, only exception being Asp222* coming from the neighboring monomer. The Fo-Fc electron density map contoured at 3σ is colored light blue.

The O2 of β phosphate is hydrogen bonded to the amide nitrogen of Met289. The thiazole ring also forms several hydrogen bonding interactions in the active site. The carboxylate moiety forms two hydrogen bonds with Arg299. The O2 of the carboxylate group is also hydrogen bonded to His237. The nitrogen atom of the

thiazole is hydrogen bonded to the carbonyl oxygen of Gly301. There is a triad formed by Met289, Met300 and Met307 on one side of the thiazole ring whereas the other side is positioned near Asp222 coming from a different monomer. Cys220, a conserved residue among THI4/THI1 orthologues, is found in a loop near the active site and did not have any electron density for the sulfur atom. This observation suggests that Cys220 can act as the sulfur source of the biosynthetic reaction (Figure 3.2).

HPLC Analysis of ZmTHI1 Bound Metabolites. The wild type and mutant proteins over-expressed in LB were analyzed for bound metabolites using HPLC at 260 nm (Figure 3.5). Co-migration with a standard sample of ScTHI4 was used for the identification of the peaks. Three peaks were obtained for the wild type enzyme corresponding to the intermediates C, B and the final product A of the pathway. The final product A concentration was found to be the most (based on the peak area), suggesting a quick conversion of the intermediates to the final product. The mutant enzyme did not show any of the adenylated metabolites bound to it suggesting that the enzyme is inactive.

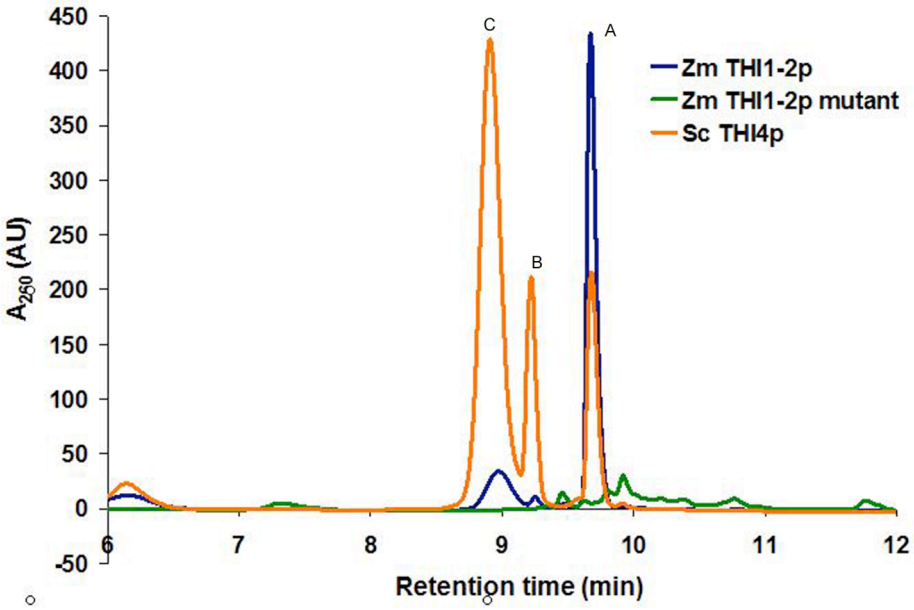


Figure 3.5. HPLC analysis for bound metabolites purifying with ZmTHI1. All traces represent the absorption at 260 nm, detected by a diode-array detector. Blue trace represents wild-type ZmTHI1 sample whereas the green and orange traces represent mutant-ZmTHI1 and ScTHI4 samples respectively.

HPLC Analysis of the Reconstitution of Activity. The biochemical assays for reconstitution of activity showed that the wild type enzyme is active (Figure 3.6a, b). The formation of the keto-imine intermediate (Figure 3.2, C) was observed in presence of NAD and the wild-type enzyme. However such reconstitution failed in case of the mutant enzyme.

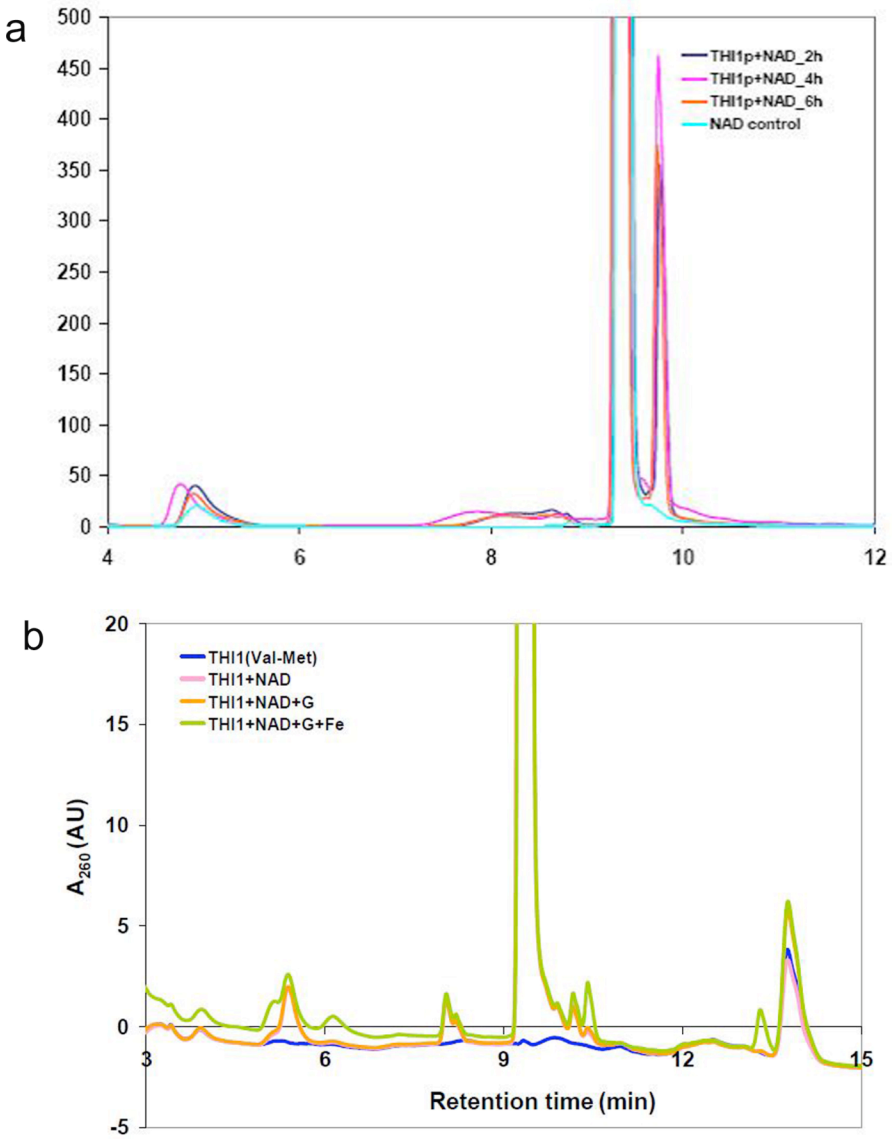


Figure 3.6. Biochemical assays for reconstitution of enzyme activity

Dynamic Light Scattering Studies. The Zetasizer software was used to extract the information of the size distribution from the measured correlation function. The size distribution curves (Figure 3.7) have the distribution of size classes in the X axis and relative intensity of the scattered light in Y axis and are called intensity distributions. The average diameter was measured at all three concentrations for each sample (Table 3.3). The molecular mass was estimated from the particle size, using an available standard curve. The calculated molecular mass indicates that the wild type enzyme exists as an octamer in solution. The molecular mass of the mutant predicted formation of an assembly larger than an octamer.

Table 3.3. Dynamic Light Scattering.

Sample[a]	d (nm)[b]	Measured Molecular Weight (kDa)	Oligomeric State[c]
ZmTHI1	12.08	304	octamer
V211M	20.33	674	-
V211A	13.62	376	octamer
V211L	11.94	302	octamer
V211F	15.76	472	-

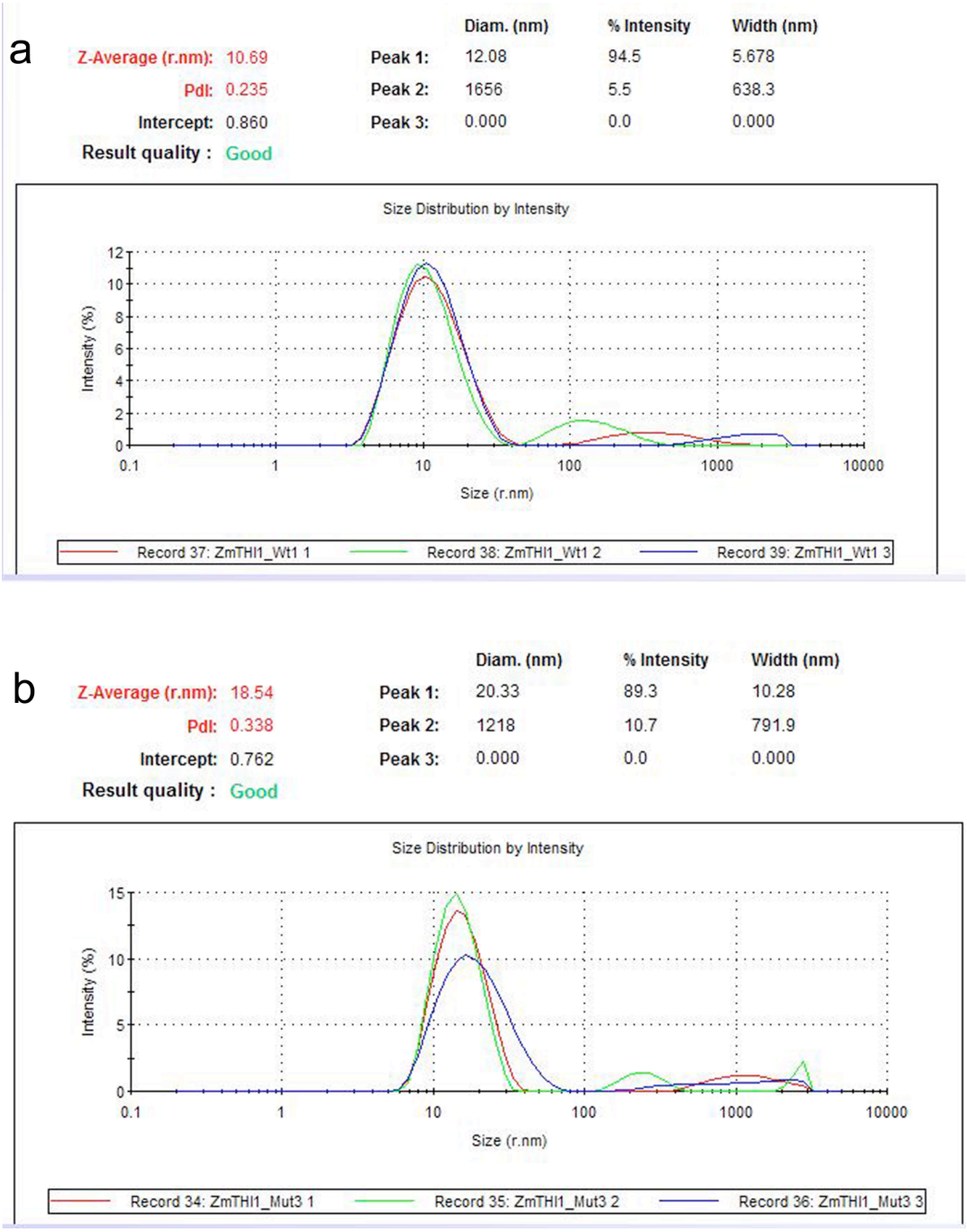


Figure 3.7. DLS studies on wild type and mutant ZmTHI1.

Site Directed Mutagenesis of ZmTHI1. To gain further knowledge about the role of Val211 in thiazole formation, several other mutants were designed. Val211 was mutated to Ala (small side chain), Leu (medium side chain), and Phe (large side chain) to investigate the steric effect of this particular residue. The biochemical assays with the mutants show that activity is dependent on the size of the substituent residue at position 211. The profile showing the bound metabolites in the protein active site suggest that the V211F mutant is least active whereas the V211A mutant shows comparable activity to the wild type enzyme (Figure 3.8).

DLS studies on the mutants were also performed under the same conditions. The V211A mutant forms an octamer whereas the other two showed larger hydrodynamic redius indicating the formation of a larger oligomeric assembly (Figure 3.9).

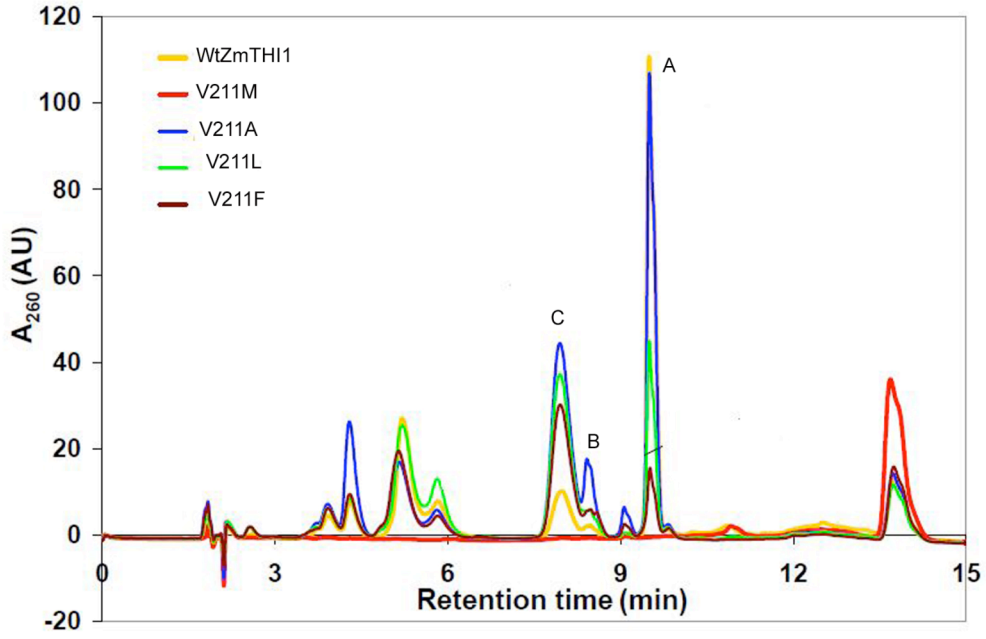
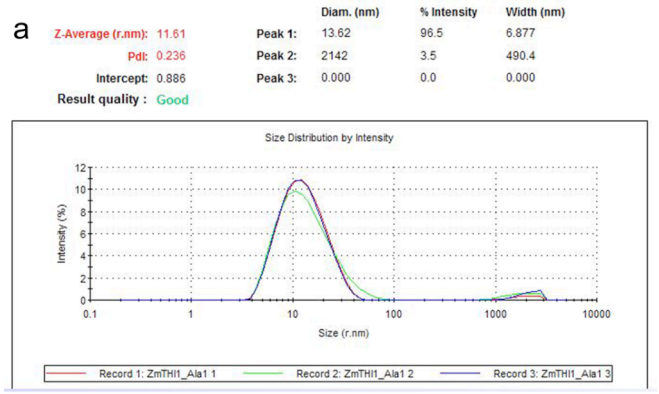
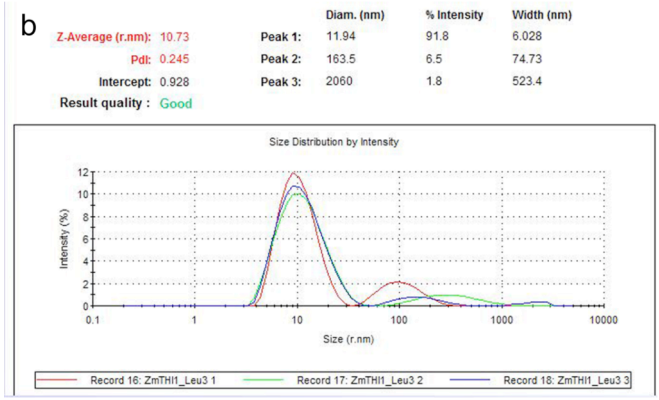


Figure 3.8. HPLC analyses for bound metabolites purifying with wild type and mutant ZmTHI1. All traces represent the absorption at 260 nm, detected by a diode-array detector. Yellow trace represents wild-type ZmTHI1, blue trace V211A, green trace V211L, brown trace V211F and red trace represents V211M sample.





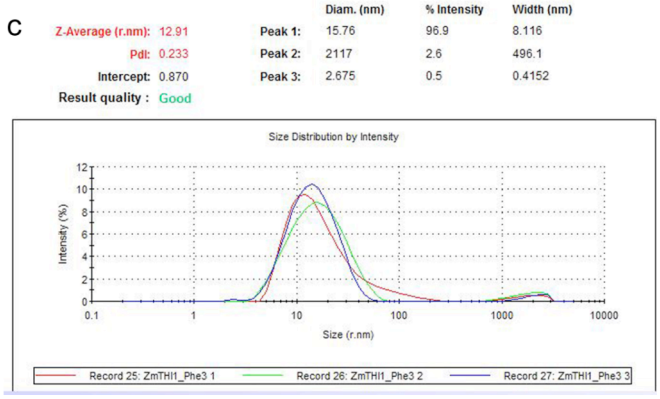


Figure 3.9. DLS studies on V211A, V211L and V211F mutants of ZmTHI1.

Section 3.4 Discussion

Comparison with AtTHI1 and ScTHI4. The structure of ZmTHI1 is the third example of thiazole synthase in eukaryotic thiamin thiazole biosynthetic pathway. The overall octameric structure (viewed as tetramer of dimer) is very similar two the other orthologues; AtTHI1 and ScTHI4, with the interfaces being mostly conserved. The monomeric fold is also very similar with significant conservation of secondary structural elements (Figure 3.10a). The RMSD between ZmTHI1 and AtTHI1 is 0.3 Å, based on alignment of 278 residues, calculated by DALI server (*30*) whereas the RMSD between ZmTHI1 and ScTHI4 is 1.3 Å, based on alignment of 272 residues. In all three structures the N-terminal and C-terminal tails are disordered. A few subtle structural differences were observed between the plant and fungal thiazole synthases, mostly involving the loop regions. An insertion of about 15 mostly hydrophilic residues (212-230), in ScTHI4 in the loop connecting β8 and β9. However sequence alignment reveals that this loop is absent in most of the THI4/THI1 orthologues and probably does not play any functional role. The other differences are observed in the loop connecting β6 and β7 (residues 177-185) and the loop connecting α7 and β10 (residues 279-285), both being longer in ScTHI4.

The active sites contain strictly conserved residues around the ligand (Table 3.4). However the thiazole conformation is different in AtTHI1, with the thiazole ring being rotated by approximately 170 ° along the C5-C6 bond (Figure 3.10b). In ZmTHI1 and ScTHI4, the nitrogen of the thiazole ring forms a hydrogen bond with Gly301 and Gly303 respectively whereas in AtTHI1, the ring undergoes flipping, resulting in binding of the nitrogen to His189, Asp190 and a Zn^{2+} ion. The electron density of the thiazole ring is poor compared to the other part of the ligand. This might suggest the possibility of a mixture of different metabolites in the active site.

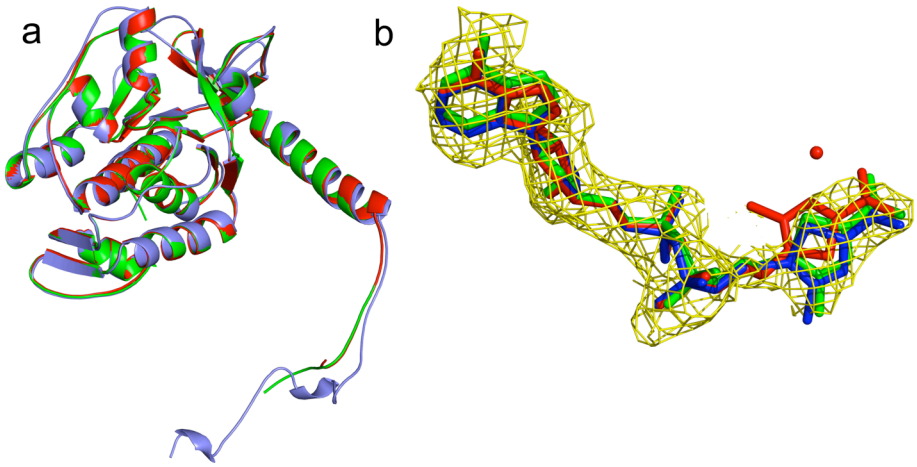


Figure 3.10. Comparison of ZmTHI1, AtTHI1 and ScTHI4. (a) Cartoon representation of the superposition of the monomers of ZmTHI1 (green), AtTHI1 (red) and ScTHI4 (blue). (b) Superposition of ADT bound in the active site of ZmTHI1 (green), AtTHI1 (red) and ScTHI4 (blue). The Zn^{2+} is colored in red. The Fo-Fc electron density map contoured at 3σ is colored light blue.

Table 3.4. Comparison of the active site residues of ZmTHI1, AtTHI1 and ScTHI4.

ZmTHI1	AtTHI1	ScTHI4
Phe241	Phe193	Phe241
Val191	Val143	Val170
Glu118	Glu70	Glu97
Gly126	Gly78	Gly105
Ala98	Cys50	Ser76
Cys220	Cys172	Cys205
Ser219	Ser171	Cys204
Met289	Met241	Met291
Met300	Met252	Met302
Met307	Met259	Met309
Arg299	Arg251	Arg301
Gly301	Gly253	Gly303
His237	His189	His237

Metal Binding Site. The active site of AtTHI1 shows the electron density for Zn^{2+} ion. The metal is octahedrally coordinated to the nitrogen of thiazole ring, His189, Asp190 and two water molecules. The side chains of corresponding residues have poor electron densities in ZmTHI1 and ScTHI4. Interestingly these two residues are highly conserved among the THI4/THI1 orthologues. However no catalytic function has been yet assigned to the Zn^{2+}. Recent studies on ScTHI4 have revealed the role of iron in full reconstitution of enzyme activity (Abhishek Chatterjee, private communication). The current reconstitution assays with iron further supports its role in thiazole biosynthesis. The absence of metal in the active site can be explained by its possible removal upon the completion of the reaction.

Mutation of Val211. The genetic mapping of *blk1* shows that the mutation V211M results in the phenotype Bladekiller 1. Val211 is a conserved residue among the THI4/THI1 orthologues. The crystal structure shows that the residue is located in a short helix, preceding the loop (215-221) (Figure 3.11), which was shown to extend into the active site of another monomer, related by a crystallographic four-fold. The Cys220, conserved among all THI4/THI1 orthologues, located in this loop is shown to be the sulfur source for thiazole formation (Abhishek Chatterjee, private communication). Although being not involved in direct interaction with the ligand, Val211 plays a crucial role in retaining the conformation of the active site loop. The replacement of the residue by larger methionine might therefore cause a conformational change in the loop preventing the thiazole formation. This explains the observation that the mutant does not purify with any of the adenylated metabolites found in case of the wild type enzyme. The failure of reconstitution using the activity assay also supports the loss of enzymatic activity caused the mutation.

The other interesting feature about Val211 is that it is located in a highly packed interface between four monomers playing an important role in the octamer

formation. It shows hydrophic interactions with Leu210 and Leu133 coming from two neighboring monomers. This leads to the fact that mutation of Val211 by a bulker side chain like Met will cause steric clashed and hinder the formation of octameric interface. In the current model Val211 was replaced by Met and all possible rotameric conformations resulted steric clashes with the surrounding residues. This is further confirmed by dynamic light scattering studies which clearly indicated octamer formation for the wild type enzyme unlike the mutant.

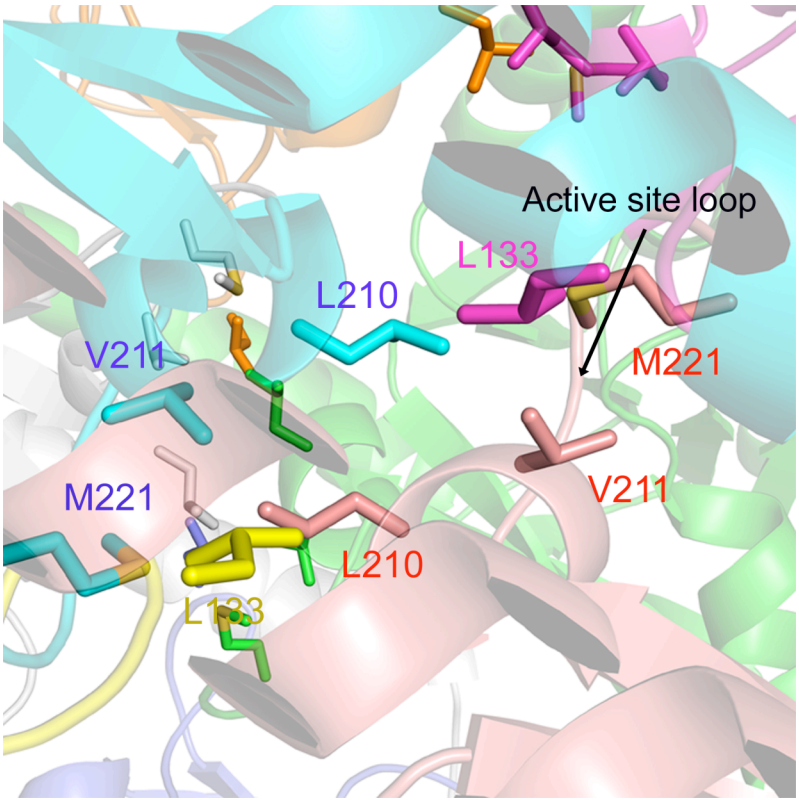


Figure 3.11. Val211 in the interface between four monomers. The interacting residues are represented in ball and stick model and color coded by monomer. The active site loop following Val211 is labeled for chain A.

Studies on Designed Mutants. To gain further insights towards the observed mutation, three mutants were designed and studied: V211A, V211L and V211F. The selection of point mutations were done based on the size of the side chains, small (Ala), medium (Leu) and large (Phe). The crystal structure of the V211A mutant revealed the octamer formation and can be described as a positive control for our hypothesis. However the active site loop (215-222) is partially disordered in this structure suggesting some structural changes by the V211A mutation. The side chain of Cys220 does not show the electron density for the sulfur atom which can be a result of loop disorder along with the thiazole formation. The difference electron density in the active site is very poor for the thiazole ring, predicting the possibility of a mixture of bound metabolites in the active site. This can be due to the slower conversion of the intermediates (B and C) to the final product because of the limited availability of Cys220 for product formation.

The profiles for the bound metabolites were analyzed for the mutants (Figure 3.6). These results show that the small side chain V211A mutant retains most of its activity but has higher amount of intermediates (B, C) bound suggesting a slower conversion to the products as compared to the wild type enzyme. The biochemical profile is concomitant with the electron density in the active site. The medium side chain V211L mutant shows a partially active profile with equal concentrations for all three adenylated metabolites (A, B and C). The large side chain V211F mutant shows very little activity exhibited by the meager amount of metabolites present. The DLS studies also show that octamer formation is hindered by substitution with larger residues at position 211 (Figure 3.7). These results definitely suggest that mutation of Val211 by a larger residue disrupts thiazole biosynthesis. These observations support the hypothesis that the mutation of valine to larger methionine results disruption of the octameric state followed by the loss of thiazole synthase activity.

Section 2.5. Acknowledgements

The collaborative studies in this chapter describe the structural and biochemical characterization of the thiazole biosynthetic enzyme ZmTHI1 and the role of Val211 towards the structure-function relationship. The project was initiated with the finding by John Woodward in Prof. Michael Scanlon's research group that a mutation in the thiazole biosynthetic gene *thi1_2* in *Zea maize* results a phenotype called Bladekiller1 (*blk1*) exhibiting premature termination meristems and poor development of leaf blades. The biochemical analysis part was done in Prof. Tadhg Begley's laboratory by Dinuka Abeydeera.

APPENDIX

Supplemental Table 3.1: Primers used in cloning of ZmTHI1

Protein	Primer sequence (5′→3′)
Wild type ZmTHI1 - F	CATATGTCCATGTCCTCCcCCAACCCGCCCTACG
Wild type ZmTHI1 - R	CTCGAGttaGGCGTCCACGACCTCGTCGTCC
Mut ZmTHI1 - F	CATATGTCCATGTCCTCCcCCAACCCGCCCTACG
Mut ZmTHI1 - R	CTCGAGttaGGCGTCCACGACCTCGTCGTCC
V211A - F	GGACATCGGCATGATCAGCGCgGcGCCCGGGATGAAGGCGCT CG
V211A - R	CGAGCGCCTTCATCCCGGGCgCcGCGCTGATCATGCCGATGTC C
V211L - F	GGACATCGGCATGATCAGCGCgcTGCCCGGGATGAAGGCGC
V211L - R	GCGCCTTCATCCCGGGCAgcGCGCTGATCATGCCGATGTCC
V211F - F	GGACATCGGCATGATCAGCGCgaTcCCCGGGATGAAGGCGC
V211F - F	GCGCCTTCATCCCGGGgAtcGCGCTGATCATGCCGATGTCC

REFERENCES

1. Begley, T. P., Downs, D. M., Ealick, S. E., McLafferty, F. W., Van Loon, A. P., Taylor, S., Campobasso, N., Chiu, H. J., Kinsland, C., Reddick, J. J., and Xi, J. Thiamin biosynthesis in prokaryotes, *Arch. Microbiol. 171*, 293-300.
2. Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R., and Patel, D. J. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch, *Nature 441*, 1167-1171.
3. Machado, C. R., Praekelt, U. M., de Oliveira, R. C., Barbosa, A. C., Byrne, K. L., Meacock, P. A., and Menck, C. F. Dual role for the yeast THI4 gene in thiamine biosynthesis and DNA damage tolerance, *J. Mol. Biol. 273*, 114-121.
4. Medina-Silva, R., Barros, M. P., Galhardo, R. S., Netto, L. E., Colepicolo, P., and Menck, C. F. Heat stress promotes mitochondrial instability and oxidative responses in yeast deficient in thiazole biosynthesis, *Res. Microbiol. 157*, 275-281.
5. Bettendorff, L. A non-cofactor role of thiamine derivatives in excitable cells?, *Arch. Physiol. Biochem. 104*, 745-751.
6. Coelho, L. S., Hueb, J. C., Minicucci, M. F., Azevedo, P. S., Paiva, S. A., and Zornoff, L. A. (2008) Thiamin deficiency as a cause of reversible cor pulmonale, *Arq Bras Cardiol 91*, e7-9.
7. Belanger, F. C., Leustek, T., Chu, B., and Kriz, A. L. Evidence for the thiamine biosynthetic pathway in higher-plant plastids and its developmental regulation, *Plant. Mol. Biol. 29*, 809-821.
8. Xiang, S., Usunow, G., Lange, G., Busch, M., and Tong, L. Crystal structure of 1-deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis, *J. Biol. Chem. 282*, 2676-2682.

9. Taylor, S. V., Kelleher, N. L., Kinsland, C., Chiu, H. J., Costello, C. A., Backstrom, A. D., McLafferty, F. W., and Begley, T. P. Thiamin biosynthesis in *Escherichia coli*. Identification of this thiocarboxylate as the immediate sulfur donor in the thiazole formation, *J. Biol. Chem. 273*, 16555-16560.

10. Settembre, E. C., Dorrestein, P. C., Park, J. H., Augustine, A. M., Begley, T. P., and Ealick, S. E. Structural and mechanistic studies on ThiO, a glycine oxidase essential for thiamin biosynthesis in *Bacillus subtilis*, *Biochemistry 42*, 2971-2981.

11. Leonardi, R., Fairhurst, S. A., Kriek, M., Lowe, D. J., and Roach, P. L. Thiamine biosynthesis in *Escherichia coli*: isolation and initial characterisation of the ThiGH complex, *FEBS. Lett. 539*, 95-99.

12. Park, J. H., Dorrestein, P. C., Zhai, H., Kinsland, C., McLafferty, F. W., and Begley, T. P. Biosynthesis of the thiazole moiety of thiamin pyrophosphate (vitamin B1), *Biochemistry 42*, 12430-12438.

13. Toms, A. V., Haas, A. L., Park, J. H., Begley, T. P., and Ealick, S. E. Structural characterization of the regulatory proteins TenA and TenI from *Bacillus subtilis* and identification of TenA as a thiaminase II, *Biochemistry 44*, 2319-2329.

14. Praekelt, U. M., Byrne, K. L., and Meacock, P. A. Regulation of THI4 (MOL1), a thiamine-biosynthetic gene of *Saccharomyces cerevisiae*, *Yeast 10*, 481-490.

15. Chatterjee, A., Jurgenson, C. T., Schroeder, F. C., Ealick, S. E., and Begley, T. P. Thiamin biosynthesis in eukaryotes: characterization of the enzyme-bound product of thiazole synthase from *Saccharomyces cerevisiae* and its implications in thiazole biosynthesis, *J. Am. Chem. Soc. 128*, 7158-7159.

16. Godoi, P. H., Galhardo, R. S., Luche, D. D., Van Sluys, M. A., Menck, C. F.,

and Oliva, G. Structure of the thiazole biosynthetic enzyme THI1 from *Arabidopsis thaliana*, *J. Biol. Chem. 281*, 30957-30966.

17. Chatterjee, A., Schroeder, F. C., Jurgenson, C. T., Ealick, S. E., and Begley, T. P. Biosynthesis of the thiamin-thiazole in eukaryotes: identification of a thiazole tautomer intermediate, *J. Am. Chem. Soc. 130*, 11394-11398.
18. Faou, P., and Tropschug, M. A novel binding protein for a member of CyP40-type Cyclophilins: N.crassa CyPBP37, a growth and thiamine regulated protein homolog to yeast Thi4p, *J. Mol. Biol. 333*, 831-844.
19. Ausubel, F. M., and Brent, F. (1987) *Current Protocols in Molecular Biology*, John Wiley and Sons, New York.
20. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Vol. 3, Cold Spring Harbor Laboratory Press, Plainview, New York.
21. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.
22. Matthews, B. W. (1968) Solvent content of protein crystals, *J. Mol. Biol. 33*, 491-497.
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool, *J. Mol. Biol. 215*, 403-410.
24. Lorentzen, E., Pohl, E., Zwart, P., Stark, A., Russell, R. B., Knura, T., Hensel, R., and Siebers, B. (2003) Crystal structure of an archaeal class I aldolase and the evolution of $(ba)_8$ barrel proteins, *J. Biol. Chem. 278*, 47253-47260.
25. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.
26. van Aalten, D. M., Bywater, R., Findlay, J. B., Hendlich, M., Hooft, R. W., and Vriend, G. (1996) PRODRG, a program for generating molecular

topologies and unique molecular descriptors from coordinates of small molecules, *J. Comput. Aided Mol. Des. 10*, 255-262.

27. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination, *Acta Crystallogr. D 58*, 1948-1954.

28. DeLano, W. L. (2002), The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA.

29. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol. 372*, 774-797.

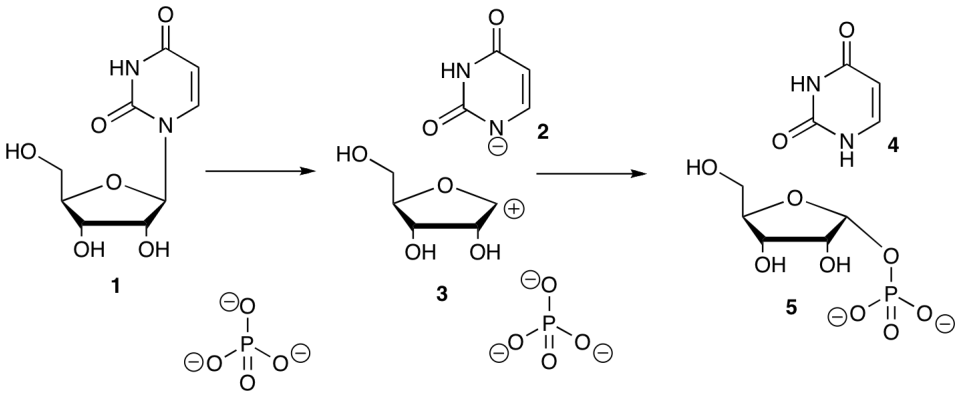
30. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol. 233*, 123-138.

CHAPTER 4

GLYCAL FORMATION IN CRYSTALS OF URIDINE PHOSPHORYLASE[1]

Section 4.1 Introduction

Uridine phosphorylase (UP; EC2.4.2.3) is a key enzyme in the pyrimidine salvage pathway, which when appropriate precursors are available provides an alternative to the energy-costly *de novo* biosynthetic pathway. UP catalyzes the reversible phosphorolysis of uridine (Urd) and 2′-deoxyuridine (dUrd), as well as their analogues to the respective nucleobases and ribose 1-phosphate (Scheme 4.1). UP is present in most organisms including prokaryotes, yeast and higher organisms including mammals.



Scheme 1

The three-dimensional structure of *Escherichia coli* UP (EcUP) has been previously studied in its unliganded form and in several complex forms (*1-4*). The structure of *Salmonella typhimurium* UP is very similar to that of EcUP (*5*). The bacterial enzyme exists as a hexamer comprised of six identical subunits. The

[1] Reproduced with permission from Paul, D., O'Leary, S., Kanagalaghatta, R., Bu, W., Toms, A., Settembre, E., Sanders, J., Begley, T. P and Ealick, S.E. (2010) *Biochemistry 49*, 3949-3509. Copyright 2010 American Chemical Society

protomer shares a common α/β fold with the nucleoside phosphorylase-I (NP-I) superfamily (*6*). Other members of this superfamily include mammalian (*7, 8*) and bacterial (*9*) purine nucleoside phosphorylases (PNPs), mammalian and bacterial methylthioadenosine phosphorylases (*10*), *S*-adenosylhomocysteine/methylthioadenosine nucleosidase (*11*) and AMP nucleosidase (*12*). Despite the differences in amino acid sequence, oligomeric state and substrate specificity, the enzymes of this superfamily are characterized by a common topology.

In mammalian cells two isoforms of UP have been identified (*13*). The class 1 isozyme is more widely distributed and has been studied in detail. The class 2 enzyme shows a broader substrate specificity and substantial differences in tissue specific expression. Recently the structure of class 1 human UP (hUP) was reported in unliganded and ligand-bound forms (*14*). The human enzyme exists as a dimer with the same structural fold as the bacterial enzyme. Nearly all active site residues are conserved between the mammalian and bacterial homologues and show similar protein-ligand interactions. Structural analogy of UP to the closely related PNP, which has been studied extensively by isotope effects, theoretical methods and inhibition (*15-17*), suggests that the chemical reaction proceeds via an oxocarbenium ion-like transition state (**3**, Scheme 4.1).

Carbenium ions have fascinated organic chemists for almost a century and the properties of this important class of reactive intermediates have been studied extensively (*18*). In addition to the nucleoside phosphorylase reaction, carbenium ions have been proposed as intermediates in several enzyme-catalyzed reactions, including the prenyl transfer and cyclization reactions involved in terpene biosynthesis (*19*), the pyrrole tetramerization involved in porphyrin biosynthesis (*20*), the glycosyl transfer reactions involved in a wide variety of glycosylations (*21*), and the thiazole/pyrimidine coupling reaction involved in thiamin phosphate formation (*22,*

23). With the exception of the latter reaction, evidence for enzyme-bound carbenium ions is always indirect, relying on substituent and isotope effects, on intermediate trapping and on inhibition studies. A possible exception comes from studies on human CD38 (*24*).

Here we present the crystal structures of EcUP treated *prior to crystallization* with 5-fluorouridine (FUrd) and sulfate, and bovine UP (bUP) treated with 5-fluoro-2′-deoxyuridine (FdUrd) or Urd, and sulfate. Unexpectedly, the electron density for each structure shows three separate species corresponding to Ura or 5-fluorouracil (FUra), sulfate and a ribosyl species, which was modeled as a UP-bound glycal. Our studies also include the determination of the structure of unliganded bUP and bUP treated with dUrd and sulfate. In the latter structure the dUrd is not cleaved.

Section 4.2 Experimental

Chemical Reagents. M9 minimal medium, L-amino acids, magnesium sulfate, calcium chloride, uracil (Ura), FUra, Urd, FUrd, dUrd and FdUrd were purchased from Sigma-Aldrich. Ni-NTA resin was obtained from Qiagen (Valencia, CA), MEM vitamin solution from Invitrogen, ampicillin, MES, ferrous sulfate, and L-selenomethionine (SeMet) from Acros and glucose from Fisher

Crystallization of EcUP. EcUP was expressed and purified as previously described (*1*). EcUP was crystallized by the hanging drop method using 2 μL of protein solution plus 1 μL of well solution. The protein crystallized under conditions very similar to those published by Burling *et al (2)*. The crystals grew best at 22 °C with the optimized mother liquor conditions of 5% - 7% PEG 4K, 0.1 M 2-(N-morpholino)ethanesulfonic acid (MES) pH 6.2 - pH 6.3, and 5% glycerol. Crystals grew over 2 days to a maximum size of ~300 μm x 300 μm x 200 μm. For complex crystallization, the protein was incubated with 10 mM FUrd and 10 mM ammonium sulfate at 4 °C for 12 h *prior to crystallization*. Co-crystals grew within two days,

under similar conditions as the unliganded form. These crystals were rod-like with maximum dimensions of 25 μm x 25 μm x 500 μm.

Cloning of bUP. The bUP gene was amplified from the cDNA clone Mbov3 139P15 (BacPac, http://bacpac.chori.org/) by PCR with the following primer pair: 5′-CTG TAC TTC CAG GGT ATG GCC TCC ACC GCG GCT GAA ACG GAG-3′ and 5′ -GAA AGC TGG GTG TCA GGC CTG CAT CAG GCG CTT C-3′. The PCR product was then used as a template for another round of thermocycling with the following primers designed to add AttB sites to the ends of the encoded protein: 5′ -G GGG ACA AGT TTG TAC AAA AAA GCA GGC TCC GGA AAT CTG TAC TTC CAG-3′ and 5′- GGG GAC CAC TTT GTA CAA GAA AGC TGG GTG-3′. The PCR product was purified and used in a Gateway BP recombination reaction with the Invitrogen vector pDONR221 following the manufacturer's instructions. Clones were screened by restriction digest and verified by sequencing. A correct clone was used in a Gateway LR recombination reaction with the plasmid pDESTF1, which is a gateway adapted vector based on the pET-system from Novagen. The plasmid pDESTF1 encodes an N-terminal 6xHis tag and is under the control of T7*lac* promoter, which is a hybrid promotor having a T7 promotor sequence followed by a *lac* operator sequence. Again, clones were screened by restriction digest. A correct clone was named pBovUP.XF1.

Expression and Purification of bUP. The plasmid pBovUP.XF1 was transformed into BL21Star(DE3) *Escherichia coli*. An overnight culture of 10 mL was grown in Luria-Bertani media at 37 ºC supplemented with 100 μg/mL ampicillin and then diluted into 1 L of medium containing 100 μg/mL ampicillin. The cells were grown at 37 ºC until they reached an OD_{600} of 0.4, and the temperature was then reduced to 20 ºC. The cells were induced with 1 mM isopropyl 1-β-D-galactopyranoside (IPTG) when the OD_{600} reached 0.6 and allowed to grow overnight

at 15 ºC. Cells were harvested by centrifugation at 9000*g* for 10 min at 4 ºC and stored at -20 ºC.

The frozen cell pellet was resuspended in 50 mL of lysate buffer (50 mM tris(hydroxymethyl)aminomethane (Tris), pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM dithiothreitol (DTT)), and cells were subjected to sonication. Cell debris was removed by centrifugation at 35000*g* and the clear lysate was loaded onto a Ni-NTA column pre-equilibrated with the lysate buffer. After the lysate was loaded, the column was washed with wash buffer (50 mM Tris, pH 8.0, 300 mM NaCl, 30 mM imidazole, 1 mM DTT) for about an hour. Pure protein was eluted with 50 mM Tris, pH 8.0, 300 mM NaCl, 250 mM imidazole and 1 mM DTT. The recombinant protein was subjected to tobacco etch virus (TEV) protease digestion, to remove the polyhistidine tag, in a buffer containing 50 mM Tris, pH 8.0, 150 mM NaCl, 1 mM DTT) for 12 h. After the completion of proteolysis, as judged by SDS gel analysis, the protein was re-loaded onto a Ni-NTA column and the flow through was collected. The non-binding material was then buffer exchanged into 20 mM Tris, pH 8.0, 150 mM NaCl. The protein was then concentrated to 3 mg/mL using an Amicon Ultra centrifugal device with a 30 kDa MWCO (Millipore). The protein was estimated to be greater than 95% pure based on SDS gel analysis. The pure protein was flash frozen and stored at -80 ºC.

Expression and Purification of SeMet Bovine UP. The expression of SeMet incorporated protein involves some modifications to the native protein expression protocol. The plasmid pBovUP.XF1 was transformed into methionine-auxotrophic B834(DE3) *E. coli*. A 50 mL starter culture in M9 minimal medium, containing 40 mg/mL L-amino acids, 0.4% glucose, 2 mM magnesium sulfate, 1% MEM vitamin solution, 0.1 mM calcium chloride, 25 μg/mL ferrous sulfate and 100 μg/mL ampicillin was inoculated with a single colony and grown overnight at 37 ºC. The

starter culture was grown in the presence of L-methionine to give a healthy start, as selenomethionine is poisonous to cells. The starter culture was then used to inoculate 1 L M9 minimal medium having the same composition but L selenomethionine in place of L-methionine. The starter cells were washed with the M9 minimal media prior to inoculation to remove the normal methionine. Induction was done using 1 mM IPTG at OD_{600} of 0.6 for 12-16 hours at 15 ºC. Cells were harvested by centrifugation at 9000*g*.

The purification process was the same as that for the native protein. The pure protein was stored in 20 mM Tris, pH 8.0, 150 mM NaCl and 1 mM DTT.

For the native protein, 1 L of cell cuture produced about 20 mg of pure protein, while for the SeMet protein, only ~4 mg of pure protein was obtained from 1 L of culture.

Crystallization of bUP. Initial crystallization conditions of unliganded bUP were determined using sparse matrix screens Crystal Screen 1 and 2 (Hampton research) at 22 ºC using the hanging-drop vapor diffusion technique. Several conditions yielded thin needle crystals that varied in quality. The optimized conditions were 100 mM MES (pH 6.3), 140 mM ammonium sulfate, 18% PEG 5K monomethylether (PEG 5K MME). The crystals grew in 3-5 days to the size of ~200 μm long rods. Drops containing 1 μL of protein and 1 μL of reservoir solution were optimal for crystal growth.

To obtain the ligand complexes, the enzyme was separately incubated with 5 mM each of Urd, dUrd or FdUrd and 5 mM ammonium sulfate for 12 h prior to crystallization. The rod-shaped 200-300 μm long crystals of the substrate complexes were obtained under very similar conditions (0.1 M MES pH 6.0-6.5, 0.2 M ammonium sulfate, 18-25% PEG 5K MME). SeMet UP crystals were grown by hanging drop vapor diffusion at 22 ºC using a well solution containing 100 mM MES

(pH 6.0), 1 M LiCl and 20% PEG 6K. In this case the needle like crystals were ~50-80 μm long and attained their maximum size in 4-6 days.

Data Collection and Processing. The data for EcUP complex were measured at beamline 8-BM at the NE-CAT beamline at the Advanced Photon Source (APS) at Argonne National Laboratory using a Quantum 315 detector (Area Detector Systems Corporation). Data were processed using DENZO and SCALEPACK (*25*). The data for native bUP, SeMet bUP and the complexes were collected at the 24-ID-C station of the NE-CAT beamline at the APS also using a Quantum 315 detector. A cryoprotectant solution of 10%-20% glycerol in the mother liquor was used to avoid damage during freezing. The data were indexed, integrated and scaled using the HKL2000 program suite (*25*). The data collection and data processing statistics are shown in Table 4.1.

Table 4.1. Data collection statistics for UP crystals. Values for the highest resolution shell are given in parentheses.

	EcUP/FUra /glycal/SO_4	SeMet bUP	bUP	bUP/FUra /glycal/SO_4	bUP/Ura /glycal/SO_4	bUP/dUrd /SO_4
Resolution (Å)	2.2	3.5	2.1	2.6	2.3	2.8
Space group	$P3_2$	$I4_122$	$P4_1$	$P4_12_12$	$P4_12_12$	$P4_12_12$
a (Å)	92.9	81.9	74.1	82.4	83.5	82.6
c (Å)	145.0	266.3	266.9	258.6	260.1	259.0
No/ASU	6	1	4	2	2	2
Matthews number ($Å^3$/Da)	2.28	3.18	2.62	3.13	3.23	3.02
Solvent content	43.3%	60%	52%	60%	61%	59.4%
Unique reflections	68027	10857	91922	28368	41493	20780
Redundancy	7.5	8.5	2.2	6.1	7.0	3.1
Completeness	99.3 (99.0)	98.9 (98.5)	95.7 (95.1)	99.5 (95.8)	100 (99.9)	89 (73.8)
R_{sym} * (%)	4.0 (9.6)	9.4 (34.4)	9.1 (36.3)	11.4 (36.4)	20.8 (54.9)	11.4(21.3)
I/s	17.9 (4.6)	11.1 (2.4)	12.4 (2.7)	20.6 (3.6)	35.0(6.14)	9.5 (2.5)

* $R_{sym} = \Sigma\Sigma_i \left| I_i - \langle I \rangle \right| / \Sigma \langle I \rangle$, where <I> is the mean intensity of the N reflections with intensities I_i and common indices *h,k,l*

Structure Determination and Refinement of the EcUP/Ura/glycal/sulfate Complex. The current structure was determined by molecular replacement using the program CNS (*26*), with the previously reported native UP monomer structure (Protein Data Bank entry 1LX7) as the search model (*2*). The refinement procedure included successive rounds of rigid body refinement, simulated annealing refinement, temperature factor refinement, and model building. A composite omit map was examined to determine the active site contents and to identify alternate conformations of side chains. Side chains and regions showing poorly defined peptide backbone electron density were then manually adjusted using the program O (*27*). Noncrystallographic symmetry (NCS) restraints improved density for some of the weaker regions. The active site ligands were clearly observed in the Fo-Fc maps, but ligand models were not included in the refinement until later steps to minimize model bias. After all protein atoms were refined, the appropriate ligand models were manually fit into the active site, followed by a few cycles of refinement. The model was examined using CNS and PROCHECK (*28*). The final refinement statistics are summarized in Table 4.2.

Table 4.2. Refinement statistics and model building for UP datasets.

	EcUP/FUra/glycal/SO$_4$	bUP	bUP/FUra/glycal/SO$_4$	bUP/Ura/glycal/SO$_4$	bUP/dUrd/SO$_4$
Resolution (Å)	2.2	2.1	2.6	2.3	2.8
Protein Atoms	10879	8862	4487	4483	4489
Water Molecules	839	980	155	334	77
Ligand Atoms	138	20	44	54	42
RMSD	0.018	0.008	0.008	0.008	0.007
bonds (Å)	1.980	1.141	1.173	1.128	1.019
angles (deg)					
R factor [a] (%)	19.7	17.9	16.7	18.0	18.2
R_{free} [b] (%)	23.9	22.5	20.5	20.8	25.0
favoured regions (%)	89.4	93.2	91.1	91.8	88.9
Additionally allowed regions (%)	10.2	6.8	8.9	8.2	11.1
Generously allowed regions (%)	0.4	0.0	0.0	0.0	0.0
Disallowed regions (%)	0.0	0.0	0.0	0.0	0.0

[a] R factor $= \sum_{hkl} \left| \left| F_{obs} \right| - k \left| F_{cal} \right| \right| / \sum_{hkl} \left| F_{obs} \right|$ where F_{obs} and F_{cal} are observed and calculated structure factors, respectively. [b] For R_{free} the sum is extended over a subset of reflections (5%) excluded from all stages of refinement.

Structure Determination and Refinement of bUP in Unliganded and Complex Forms. To determine the structure of bUP, the single wavelength anomalous diffraction (SAD) method was used. Ten Se sites out of the 12 possible sites were found by the program SHELXD (*29*). SAD phasing was done by the maximum likelihood program MLPHARE (*30*) and finally a density modified map was calculated using the program SHARP (*31*), which was used for initial model building at 3.5 Å. Manual model building was done using the interactive graphic program Coot (*32*). After a few rounds of model building, a biologically relevant crystallographic dimer in the SeMet UP unit cell was identified and used as the starting model for molecular replacement with the higher resolution native dataset. Molecular replacement was carried out, with the dimer as starting model, using CNS (*33*) and the second dimer was found by translational search after fixing the first one in the unit cell. Initial electron density maps were further improved by fourfold NCS averaging with the RAVE suit of programs (*34*). The refinement procedure involved successive rounds of rigid-body refinement, simulated-annealing refinement, temperature factor refinement and model rebuilding. Initial refinement cycles were performed using CNS and final rounds with the PHENIX suit of programs (*35*). The bUP complex structures were determined by molecular replacement using CNS with the refined native bUP-dimer as the search model. The active-site contents and alternative side chain conformations were determined using F_o-F_c and composite omit electron density maps. Tight NCS restraints were applied during initial rounds of refinement to improve density in the weaker regions and were slowly relaxed during the final rounds of refinements. The parameter and topology files for the ligands were generated using the Dundee PRDRG2 server (*36*) . The sulfate ions and water molecules were addedand refined in CNS. The refinement statistics are summarized in Table 4.2.

Analytical Ultracentrifugation. Analytical ultracentrifugation was used to determine the oligomeric states of EcUP and bUP. The software package DCDT+ (*37*) was used for data analysis. The basic principle involves a time derivative approach for sedimentation velocity analysis. The sedimentation coefficient distribution function is fitted to a Gaussian curve to determine concentration, sedimentation coefficient and diffusion coefficient of a particular species. However the molecular weight determination also requires the density of the solvent, the partial specific volume of the protein, the viscosity of the solvent, and its temperature dependence in order to account for the effect of temperature and solvent on sedimentation property. The program Sedenterp (*38*) was used for the calculation of those properties.

^{1}H NMR Spectroscopy of EcUP-catalyzed FUrd Hydrolysis in the Presence and Absence of Sulfate. ^{1}H NMR spectra were obtained using a Varian INOVA 600 MHz spectrometer and a Varian H{C/N} probe equipped with triple-axis pulsed field gradients. A solution (0.5 mL) was prepared containing sodium sulfate (60 mM) and FUrd (10 mM) in 50 mM ammonium acetate buffer (prepared with D_2O, pD 7.4). After addition of UP (0.25 mL, final concentration 140 *m*M), the field-frequency lock and magnet *Z*-shims were quickly optimized. 60 spectra were then acquired at 81 s intervals, followed by a further 11 spectra acquired at 10 minute intervals. Each final spectrum was produced by averaging 16 transients. The intervals between acquisition of the spectra were determined by appropriately arraying the pre-acquisition delay. The residual solvent resonance was suppressed by presaturation. Resonances corresponding to the C6-*H* nuclei of FUrd and FUra were integrated for the individual spectra and used to determine the concentration of the substrate and product as a function of reaction time. The concentrations were calculated at the individual time points as the product of the initial substrate concentration and the integral value of

substrate or product expressed as a fraction of the total integral value for the two species; no pyrimidine side products were observed during the reaction. The identity of the FUra product was established by comparison of its ^{1}H NMR spectrum with that of a commercially available standard. FUrd and bUP were present at concentrations of 5 mM and 0.01 mM respectively.

^{1}H NMR Spectroscopy of bUP-catalyzed FUrd Hydrolysis in the Presence and Absence of Sulfate. NMR spectra were acquired on a Varian INOVA 600 MHz instrument equipped with a Varian H{C/N} inverse-detection probe and triple-axis pulsed field gradients. The residual HDO resonance was suppressed by presaturation. Immediately prior to the experiment, the enzyme was buffer-exchanged into a freshly-prepared buffer composed of 0.1 M ammonium bicarbonate and 0.15 M sodium chloride in D_2O (pD 7.1). For observation of the time course of the hydrolysis reaction, the enzyme was added to a solution of FUrd (5 mM) in the same buffer. In a second experiment, potassium sulfate was also included at a final concentration of 5 mM prior to the addition of enzyme. The final enzyme concentration was ~10 *m*M. After mixing the enzyme and substrate solutions, the instrument shims were optimized and the time course of the reaction was followed by acquiring 30 traces at intervals of three minutes over 1.5 h. Each trace was generated from an average of 32 transients. The spectra were processed by applying exponential multiplication (0.35 Hz) to the free-induction decay prior to Fourier transformation.

Figure Preparation. All figures were prepared using ChemDraw and PyMOL (*39*).

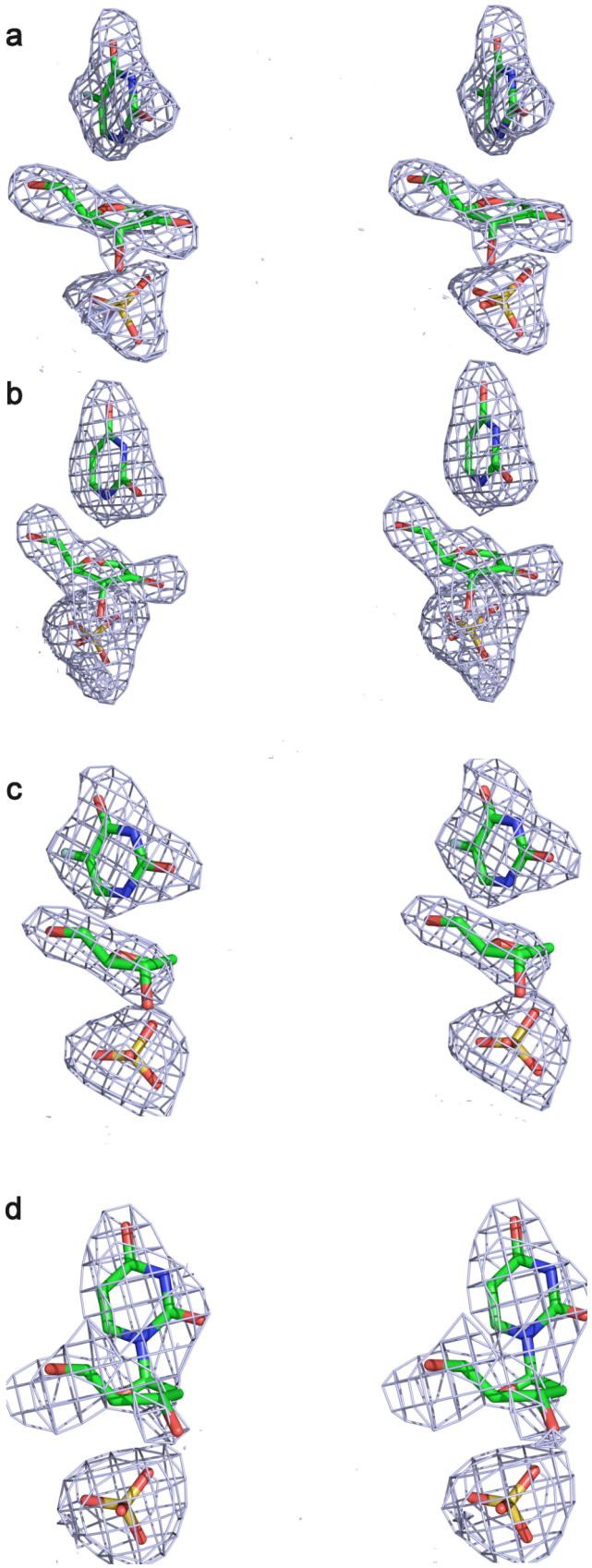
Section 4.3 Results

Observation of an Unexpected Species in the Active Sites of EcUP and bUP. The active sites of EcUP treated with FUrd and sulfate, of bUP treated with FdUrd and sulfate, and of bUP treated with Urd and sulfate showed three distinct species (Figure

4.1a,b,c). The three species were well separated and corresponded to Ura (or FUra), sulfate and a ribosyl species which we model as a glycal. The structure of bUP treated with dUrd and sulfate showed uncleaved nucleoside (Figure 4.1d).

Structure of the EcUP/FUra/Gycal/Sulfate Complex. The three dimensional structure of EcUP has been previously well characterized (Figure 4.2a,b) including the identification of key active site residues (*1-4*). The active site of the enzyme can be divided into a base, a sugar and a phosphate binding site (Figure 4.3a). In the EcUP complex, the sites are occupied by FUra, glycal and sulfate, respectively. The binding site of sulfate is nearest to the protein's surface and involves side chain interactions with Arg30, Arg91 and Thr94 from one subunit and Arg48 from the neighboring subunit. Thr94 and Gly26 are involved in phosphate binding via main-chain interactions. The 5-hydroxyl group of the glycal is stabilized by a hydrogen bonding interaction with $N^{\epsilon 2}$ of His8 coming from the other adjacent monomer and an active site water molecule. The 3-hydroxy group of the glycal is stabilized by a hydrogen bond with the side chain of Glu198. The O4 of FUra is hydrogen bonded to Arg168 and an active site water molecule that is in turn hydrogen bonded to Arg223. O2 and N3 form hydrogen bonds with the side-chain of Gln166. The FUra is also involved in a herringbone stacking interaction with Phe162. The 5-fluoro substituent of FUra is located in a hydrophobic pocket composed of residues Ile220 and Val221.

Figure 4.1. Electron density for EcUP and bUP complexes. (a) Stereoview of the active site electron density for the EcUP/FUra/glycal/sulfate complex. (b) Stereoview of the active site electron density for the bUP/Ura/glycal/sulfate complex. (c) Stereoview of the active site electron density for the bUP/FUra/glycal/sulfate complex. (d) Stereoview of the active site electron density for the bUP/dUrd/sulfate complex. All F_o-F_c electron density maps are contoured at 3σ. C atoms are colored in green, N atoms in blue, O atoms in red and S atoms in yellow.


a
b
c
d

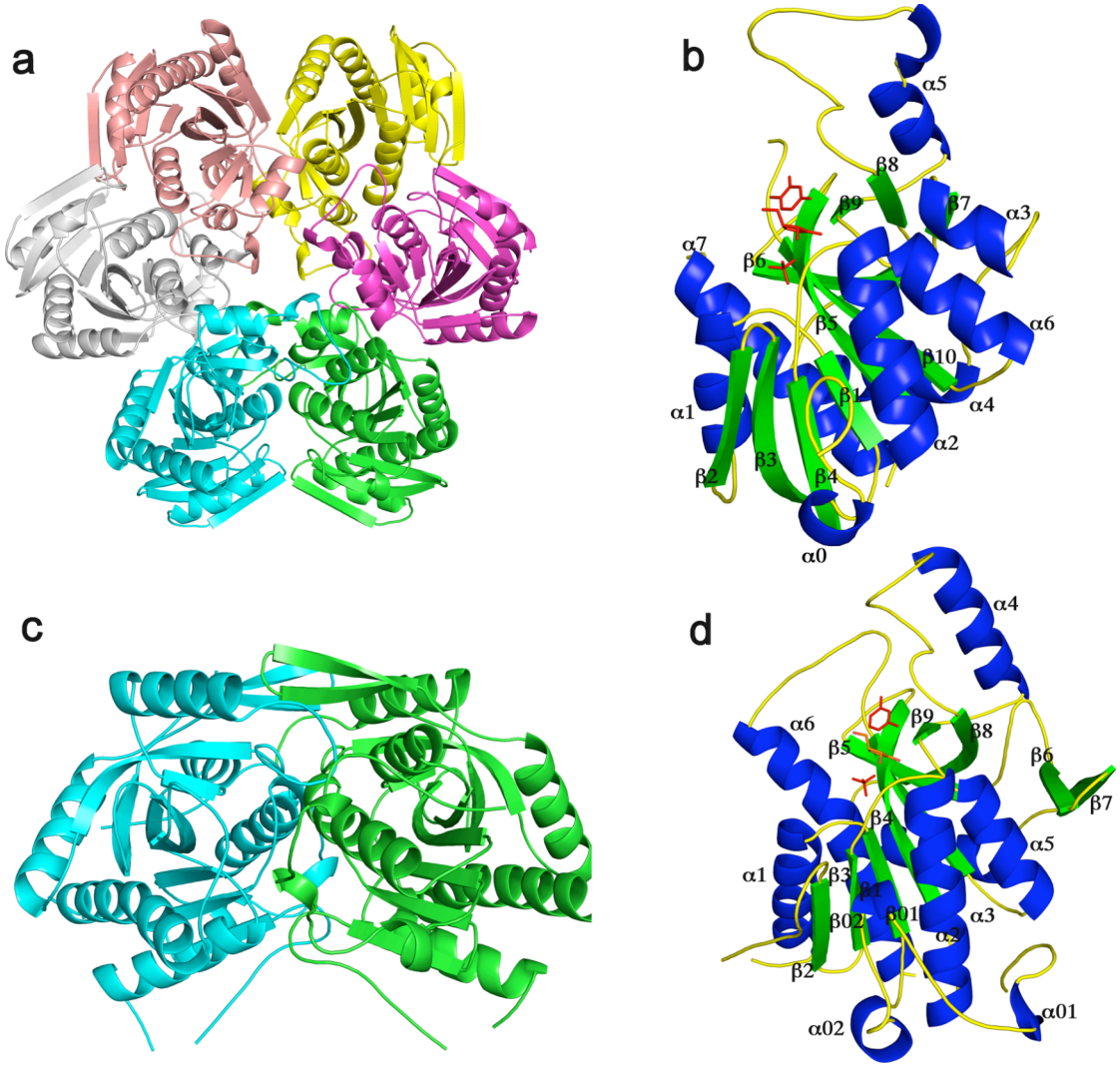


Figure4. 2. Structures of EcUP and bUP. (a) Cartoon representation of the EcUP hexamer color coded by monomer. (b) Cartoon representation of the EcUP monomer color coded by secondary structure. (c) Cartoon representation of the dimer of bovine UP color coded by monomer. (d) Cartoon representation of the monomer of bUP colored by the secondary structural elements. The α-helices are colored blue, β-strands are colored green and the loops are colored in yellow. The ligands are represented as stick models and are shown in red.

Structure of bUP. The basic fold of the nucleoside phosphorylase-I superfamily is conserved in bUP (Figure 4.2d). The characteristic feature of the superfamily is a centrally located β-sheet, which forms a distorted β-barrel, surrounded by several α-helices. The overall fold is highly similar to the recently published structure of hUP (*14*). The sequence similarity between the human and bovine enzymes is 80%. The overall three-dimensional structures are also similar, with an RMSD of 0.8 Å for 289 residues, and both enzymes are homodimers in the crystal structures. The dimer (Figure 4.2c) has a globular shape with dimensions of 65 × 40 × 25 $Å^3$. Each homodimer contains two active sites located at the dimer interface and utilizing residues from both chains. The dimer interface mostly includes the N terminal end of one chain, consisting of residues 20-24, and a loop region of the other chain, containing residues 220-224.

Structures of bUP complexes. The active site residues of bUP are very similar to those of hUP (*14*). In the current study, structures of three bUP complexes were determined: bUP/FUra/deoxyglycal/sulfate, bUP/Ura/glycal/sulfate, and bUP/dUrd/sulfate (Figure 4.3b, c, d).

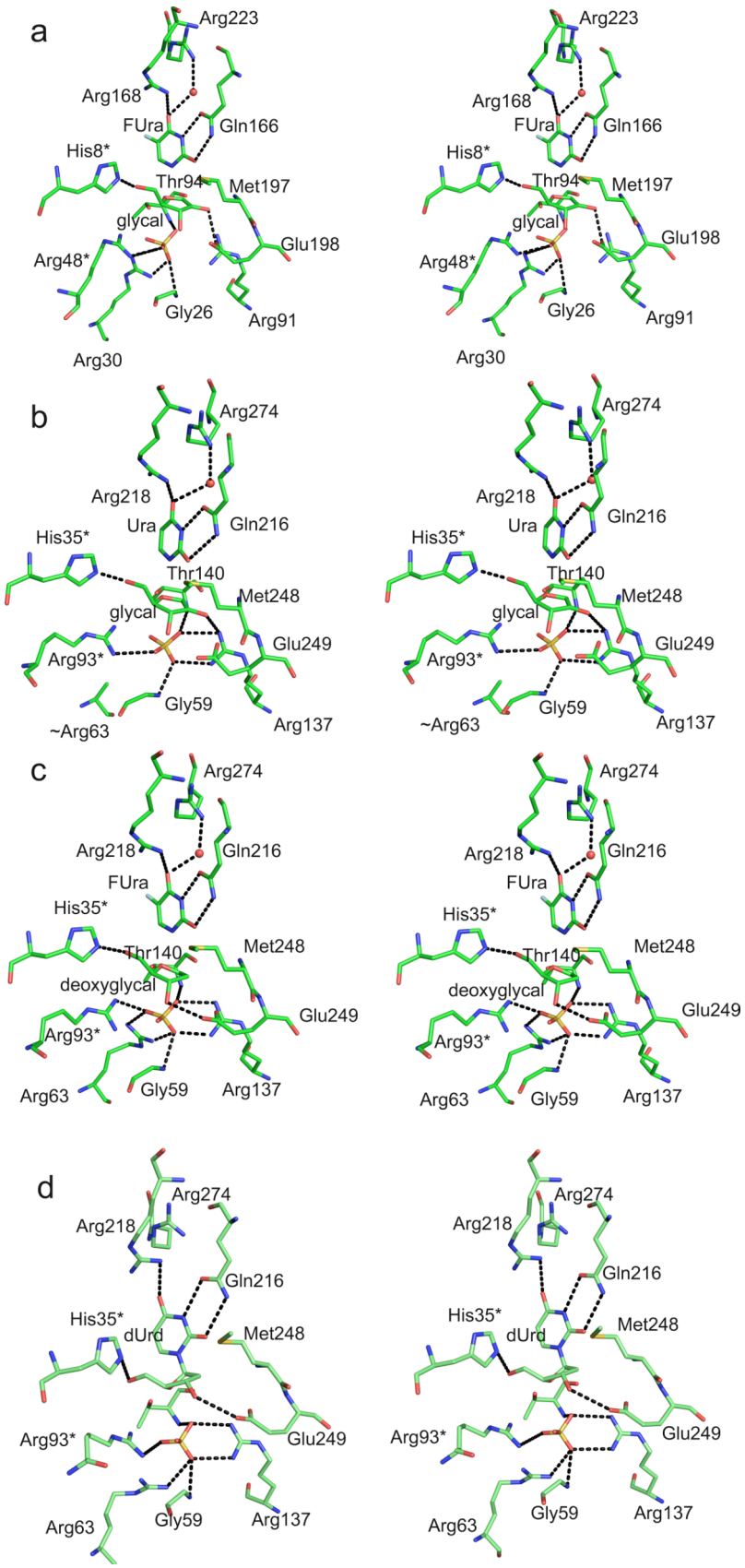
As was the case for EcUP, the sulfate binding site is more exposed to the surface than either the base or sugar binding sites. The residues involved in sulfate binding are Gly59, Arg137, and Thr140 from one subunit and Arg93 from the other. The sulfate ion forms hydrogen bonds with the side chains of Arg137 and Arg93, the amide nitrogen atom of Gly59, and both the side chain and amide nitrogen atoms of Thr140. Arg63 forms a hydrogen bond with sulfate in only the bUP/FUra/deoxyglycal/sulfate complex, but the density is very weak in the other two complexes, suggesting that the side chain is disordered.

The most interesting aspect of the current study is the observation of glycals in two of the bUP complexes. The glycal site is sandwiched between the Ura (FUra) and sulfate

binding sites in the bUP/FUra/deoxyglycal/sulfate and bUP/Ura/glycal/sulfate complexes. The side chain of Glu249 forms a hydrogen bond with the 3-hydroxyl group of the ribose ring. The 5-hydroxyl group hydrogen bonds with N^{e2} of His35 coming from the adjacent monomer. An active site water molecule also makes a hydrogen bond with the 5-hydroxyl group. The ribosyl moiety in the bUP/dUrd/sulfate complex utilizes the same interactions with the residues as the glycal.

The Ura (FUra) binding site of the enzyme is located in the cleft between two monomers and is comparatively less exposed to the surface. In the active site, the O4 atom of Ura forms a hydrogen bond with Arg218 and an active site water molecule that is in turn hydrogen bonded to Arg274. Residue Gln216 forms hydrogen bonds with O2 and N3 of the pyrimidine. Phe212 forms herringbone packing interactions with the pyrimidine. The 5-substituent of FUra inserts into a pocket formed by the side chains of Leu271 and Leu272. The side chain of Ile280 is also near this hydrophobic pocket. After glycal fomation, O2 of the Ura/FUra would be protonated. However, tautomerization would result in the keto tautomer of uracil, which is more stable at the crystallization pH (6-6.5). Consequently, the keto form was modeled into the density and the side chain orientation of Gln216 is determined accordingly; however, it is also possible that because of additional stabilizing interactions in the active site, the enzyme-bound tautomer may not be the one favored in solution.

Figure 4.3. Active sites of EcUP and bUP. (a) Stereoview of the active site of the cUP/FUra/glycal/sulfate complex. (b) Stereoview of the active site of the bUP/Ura/glycal/sulfate complex. (c) Stereoview of the active site of the bUP/FUra/glycal/sulfate complex. (d) Stereoview of the active site of the bUP/dUrd/sufate complex. Important surrounding residues are shown in stick representation. C atoms are colored in green, N atoms in blue, O atoms in red and the S atom in yellow.


a
Arg223
Arg168
FUra
Gln166
His8*
Thr94
Met197
glycal
Glu198
Arg48*
Gly26
Arg91
Arg30
b
Arg274
Arg218
Ura
Gln216
His35*
Thr140
Met248
glycal
Glu249
Arg93*
Gly59
~Arg63
Arg137
c
Arg274
Arg218
Gln216
FUra
His35*
Thr140
Met248
deoxyglycal
Glu249
Arg93*
Arg63
Gly59
Arg137
d
Arg274
Arg218
Gln216
His35*
dUrd
Met248
Glu249
Arg93*
Arg63
Gly59
Arg137

Second Sulfate Binding Site. The F_o-F_c electron density map, in the UP/Urd/sulfate complex structure shows a large peak suggesting the possibility of a second substrate binding site. This site is located near the surface and exposed to the solvent ion channel. The main interaction involves hydrogen bonds with several water molecules, along with Tyr126, His127 and His129. Because the crystallization conditions included high sulfate concentrations, this site may be an artifact of crystallization.

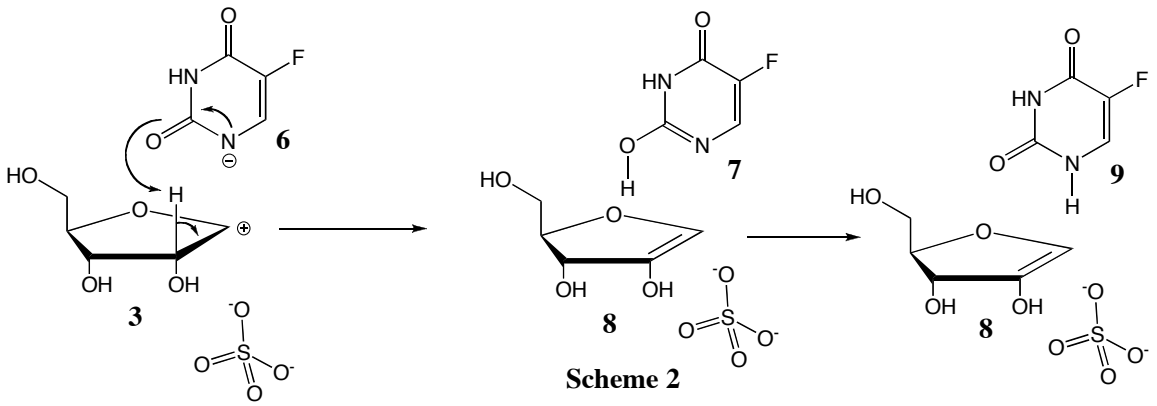
Analytical Ultracentrifugation Results. Table 3 shows the results of the data analysis. The analytical ultracentrifugation data show that bUP exists as a dimer in solution whereas the EcUP is a hexamer.

Table 4. 3. Analytical Ultracentrifugation

Sample[a]	Concentration (mg/mL)	Measured Molecular Weight (kD)	Calculated Molecular Weight (kD)	Oligomeric State
EcUP	2.0	181.89	163	hexamer
EcUP	1.0	176.26	163	hexamer
bUP	0.5	62.89	68	dimer

[a]The samples were buffer exchanged in 20 mM Tris (pH 7.0) and 150 mM NaCl.

Steady-State Reaction of the UP Enzymes with FUrd under Crystallization Conditions. Observation by ^{1}H NMR of the reactions of FUrd catalyzed by bUP and EcUP in the presence and absence of sulfate demonstrated that, under the conditions used for incubation of the enzymes with fluorinated pyrimidines and sulfate prior to crystallization, all of the nucleoside was hydrolyzed regardless of whether sulfate was present (Figure 4.4). EcUP was also capable of catalyzing the hydrolysis of Urd at a considerably slower rate than for FUrd (data not shown). We could not detect the accumulation of any products except for FUra and the four anticipated isomers of D-ribose (limit of detection ~5%). Our enzyme preparations removed phosphate from the protein and the NMR spectra showed no accumulation of ribose 1-phosphate. Furthermore, the NMR signals for the anomeric protons of the four ribose isomers were doublets, indicating that solvent-derived deuterium was not incorporated at the C2 position of the sugar on the time scale of the experiment indicating that glycal formation is irreversible under the conditions employed (Scheme 4.2).



Scheme 2

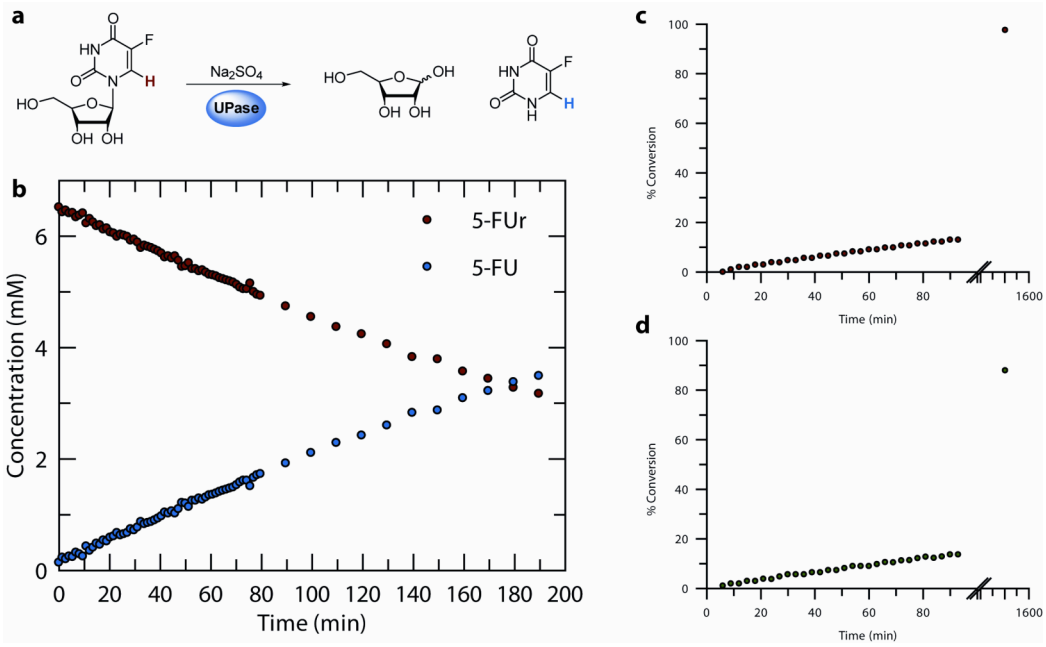


Figure 4.4. Hydrolysis of Urd in the presence of UP. (a) Reaction of FUrd occurring in the presence of UP. (b) Time course of the reaction catalyzed by EcUP and monitored by ^{1}H NMR spectroscopy. (c) Time course of hydrolysis of FUrd catalyzed by bUP in the absence of sulfate or phosphate. (d) Hydrolysis of FUrd catalyzed by bUP in the presence of potassium sulfate (5 mM) under conditions otherwise identical to (c). In the case of (c) and (d) the enzyme was present at approximately one tenth the concentration used for crystallization.

Section 4.4 Discussion

Crystal Packing for bUP. The crystal structures of unliganded, complexed and SeMet bUP show an interesting variation in crystal packing. While all of the crystals belong to the tetragonal crystal system, and have similar unit cell dimensions, three different space groups are represented (Table 2.1); however, the dimeric UP structure remains intact in all of the three space groups. In the SeMet UP (space group $I4_122$), the asymmetric unit is a monomer and the crystallographic twofold axes generate the dimers. In the UP complex structures (space group $P4_12_12$), the asymmetric unit contains one complete dimer.

The unliganded structure (space group $P4_1$) has two complete dimers in the asymmetric unit. For the lower symmetry cases, the NCS mimics the crystallographic symmetry of the highest symmetry space group.

Conformational Differences between Unliganded and Complexed bUP. The present studies report the structures of unliganded bUP, and the bUP/FUra/glycal/sulfate, bUP/Ura/glycal/sulfate, and bUP/dUrd/sulfate complexes. Comparisons of the bUP structures show conformational changes upon ligand binding (Figure 4.5). Conformational changes were also reported for hUP upon inhibitor binding (*14*). Substrate binding is associated with movement of a loop region (277-286), which forms a lid over the active site and results in a "closed" conformation for the complexes. An interesting conformational change occurs for Tyr34 where the aromatic ring rotates about 90° towards the base binding site, covering the hydrophobic pocket. In the unliganded bUP structure, Tyr34 is hydrogen bonded to Asp220 of the neighboring monomer. However, the dimeric interaction is retained by repositioning of Gln279, which makes a hydrogen bond with Asp220 after Tyr34 is displaced. The side chain of Ile280 moves towards the hydrophobic pocket, sealing the binding site, and the phenyl ring of Phe212 tilts towards the pyrimidine ring resulting in a herringbone stacking interaction. Additionally, the side-chain of Arg218 is shifted towards the active site, forming a hydrogen bond to O4 of Ura (FUra). This residue and Arg274 (which hydrogen bonds to O4 through a bridging water molecule) stabilize the negative charge that builds up in the transition state.

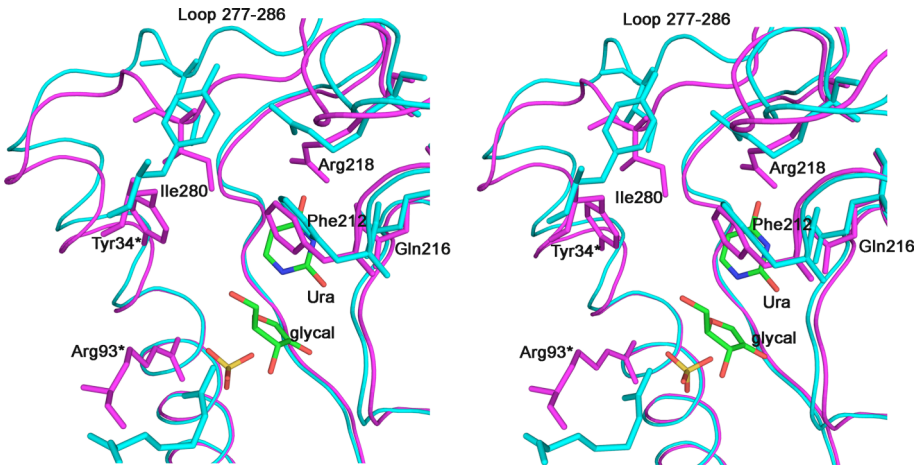


Figure 4.5. Stereodiagram showing the conformational change upon ligand binding in bUP with FUra, glycal and sulfate in the active site. The residues showing conformational changes are represented with side chains. The backbone and side chains are colored in cyan for the unliganded and in pink for the ligand-bound bUP. C atoms are colored in green, N atoms in blue, O atoms in red and S atoms in yellow for the ligand. Asterisks indicate residues from an adjacent monomer

The position of the side chain of Arg93 differs between the unliganded and complex structures. In the latter case the conformation of this residue is stabilized by a hydrogen bond to the enzyme-bound sulfate. Overall, the side chains of residues 89-93 rearrange to increase dimeric interactions in the complexed structures. The side chain of Met109 also moves towards the ribose binding site, although this residue has unusually high B-factors and its positioning may be unreliable.

Comparison of bUP and EcUP. bUP contains 309 amino acid residues per monomer whereas EcUP contains 254 amino acid residues per monomer with ~27% sequence identity between each other. Nevertheless, even though the sequence similarity is low, the basic topology is the same for the two enzymes (Figure 4.6a). The RMSD between the two structures is 2.8 Å for 236 residues, as calculated by the DALI server (*40*). An interesting aspect of the mammalian enzyme is the presence of 27 additional residues at the N terminus. The first 14 residues are not observed in the

bUP crystal structure but residues 22-52 form a β-turn-β motif, capped by 3_{10}-helices on each side (α01β01β02α02). One of the 3_{10}-helices is present in the bacterial enzyme. A number of other structural differences are seen in loop regions. The loop composed of residues 78-83 in the bovine enzyme (corresponding to residues 39-42 in EcUP), connecting α1 and β2, shows an insertion in the bovine enzyme.

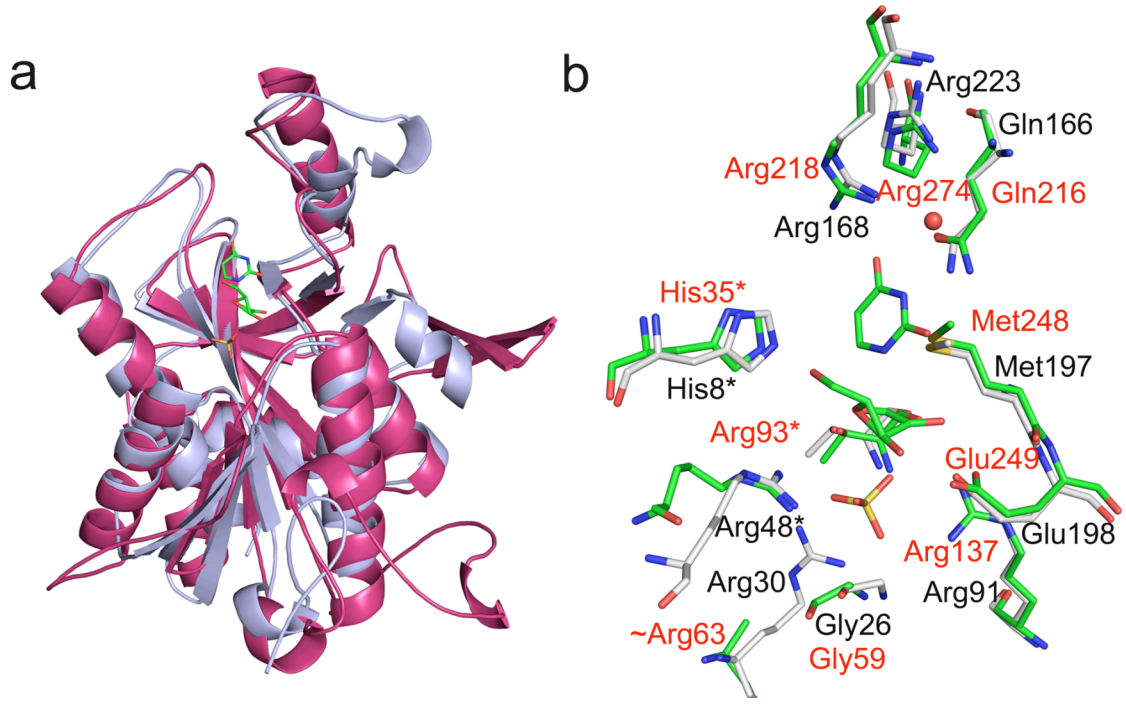


Figure 4.6. Comparison of EcUP and bUP. (a) Cartoon representation of the superposition of the monomers of bUP (pink) and EcUP (light blue). The ligands are shown in stick representation, C atoms are colored in green, N atoms in blue, O atoms in red and S atoms in yellow. (b) Superposition of active sites of bUP and EcUP with the active site showing FUra, glycal and sulfate. N atoms are colored in blue, O atoms in red, S atoms in yellow and C atoms in green for bUP and grey for EcUP. Asterisks indicate residues from an adjacent monomer.

The loop joining β2 and β3 (residues 88-95 in bUP and 45-51 in EcUP) exists in a slightly different conformation in the bUP. The loop (residues 99-102 in bUP and 55-62 in EcUP) between β3 and β4 is shorter in bUP. Another insertion occurring in the mammalian enzyme is the loop connecting α2 and β5 composed of residues 126-134 (residues 84-89 in EcUP). A structural change observed in bUP is the replacement ofa 3_{10} helix in EcUP (residues 118-124) by a β-turn-β motif composed of residues 165-178 in bUP. The turn followed by α4 exists has slightly different conformations in both the enzymes (residues 224-242 in bUP and 174-191 in EcUP). Despite the structural differences, the active sites are highly conserved between these two enzymes with very similar ligand binding residues (Table 4.4 and Figure 6b).

A significant difference observed in the crystal structures of the two enzymes is that the mammalian enzyme exists as a dimer whereas the bacterial enzyme forms a hexamer. These results were confirmed by analytical ultracentrifugation (Table 3). The bacterial hexameric assembly can be described as a trimer of dimers in which the active site is located at the dimer interface. The loss of trimerization in the mammalian enzyme can be attributed to at least two factors. The dimeric interactions in bUP are increased by additional architecture at the N terminus, offsetting the stability that would be gained from further trimerization. In particular, residues 20-24 at the N terminus interact with residues 220-224 to provide additional stabilization of the dimer. Another possible factor interfering with the trimerization is the presence of a long loop composed of residues 167-176, which replace a 3_{10} helix in EcUP.

The formation of a hexameric assembly appears in general to be a characteristic of the microbial enzymes. Interestingly, bacterial PNP exists as a hexamer organized as a trimer of dimers, and superimposes closely with the EcUP hexamer. However, the mammalian PNPs form a trimers in which the three active sites are at the trimer interface. While the protomers of hexagonal UP, dimeric UP,

hexameric PNP and trimeric PNP all have the same fold, only the latter has a fundamentally different oligomeric organization.

Glycal formation at the UP Active Site. The EcUP structure showed three clearly separated regions of density corresponding to FUra, a ribose derivative, and sulfate (Figure 1a). The C2 atom of the ribosyl species shows a tricoordinate and planar geometry. The distances between C1 of the ribose and N1 of FUra, and O1 of the sulfate are 2.93 Å and 2.87 Å, respectively, much longer than the corresponding C-N and C-O covalent bond lengths. We propose that this species is a glycal (or its keto tautomer) rather than an oxocarbenium ion because oxocarbenium ions are unlikely to be stable. Several examples of enzyme-catalyzed glycal formation have been reported. Nucleoside 2-deoxyribosyltransferase (*41, 45*) reversibly forms D-ribal from D-ribose and 2-acetamidoglucal is formed by UDP-*N*-acetylglucosamine 2-epimerase (*42, 44*). a-1,4-glucan lyase catalyzes the formation of glucal (*43*) and a glycal generated from sialic acid was observed in crystals of the influenza B virus neuraminidase (*46*).

The interactions between the putative glycal and the active site of EcUP are shown in Figure 3a. O2 of FUra is positioned directly above the C2 of the glycal at a distance of 3.82 Å in an ideal position to function as a base for proton abstraction from C2' of the substrate. The enzyme stabilizes negative charge on the C4 carbonyl group of the pyrimidine by hydrogen bonding to to Arg168 and to a water molecule (which in turn is hydrogen bonded to Arg223). Negative charge on the C2 carbonyl group is stabilized by a hydrogen bond to the amide of Gln166.

A mechanistic proposal for glycal formation is outlined in Scheme 2. In this mechanism, the C2 carbonyl oxygen of the fluorouracil anion 6 deprotonates the oxocarbenium ion 3 to give the glycal. Pyrimidine tautomerization completes the reaction. Several variants of this mechanism are also possible. For example, glycal formation could occur by a concerted syn elimination involving a late transition state

as described for a-1,4-glucan lyase . It is also possible that the oxocarbenium ion is generated from ribose and fluorouracil because most of the substrate has undergone hydrolysis before crystallization begins (Figure 4).

In the case of the bovine enzyme, both the Urd and FdUrd complex structures showed three clearly separated regions of density corresponding to the Ura (FUra), the ribosyl moiety and sulfate (Figure 1b, c). The distances between the C1 of the glycal and the N1 of the Ura or FUra are 2.54 Å or 2.53 Å, respectively. The distances between the glycal C1 and sulfate O1 atoms are 2.87 Å and 2.84 Å in the cases of the Ura/glycal/sulfate and FUra/deoxyglycal/sulfate complexes, respectively. These observations suggest formation of glycal in a similar manner as described for the bacterial enzyme. The key interactions for binding of the glycal in the active site of bUP are very similar to those in EcUP. The action of fluorouracil as a general base for deprotonation of the ribose C2 is again facilitated by the stabilization of the fluorouracil anion by hydrogen bonds to Arg218 and to a water molecule, which in turn hydrogen bonds with Arg274. The negative charge on the C2 carbonyl oxygen is stabilized by a hydrogen bond with the amide nitrogen of Gln216. The glycal is further stabilized by hydrogen bonding interactions with the side-chains of Met248 and Glu249.

In the bUP/dUrd/sulfate complex, dUrd remains uncleaved (Figure 1d) and mimics the enzyme-substrate complex. The sulfate ion occupies the phosphate binding site. The distance between the C1′ ribose and O1 of sulfate, where the new C-O bond would form in the case of phosphate, is 3.51 Å.

The trapping of the glycal was unanticipated because nothing about the room temperature steady state behavior of the enzyme predicted its trapping. We used ^{1}H NMR spectroscopy to follow the time course of reaction of FUrd with bUP and EcUP, in the presence and absence of sulfate, under conditions similar to the pre-incubation

used in the crystallization protocol. These experiments (Figure 4) show that both the mammalian and bacterial enzymes catalyze the irreversible hydrolysis of the FUrd *N*-ribosyl bond, regardless of whether sulfate is present. Hydrolysis is complete within the period of the pre-incubation. The enzyme also catalyzes hydrolysis of Urd at a much slower rate under similar conditions (data not shown). We did not observe time-dependent inhibition of the enzyme resulting from the possible accumulation of the trapped glycal at the active site.

REFERENCES

1. Bu, W., Settembre, E. C., el Kouni, M. H., and Ealick, S. E. (2005) Structural basis for inhibition of *Escherichia coli* uridine phosphorylase by 5-substituted acyclouridines, *Acta Crystallogr D 61*, 863-872.
2. Burling, F. T., Kniewel, R., Buglino, J. A., Chadha, T., Beckwith, A., and Lima, C. D. (2003) Structure of *Escherichia coli* uridine phosphorylase at 2.0 Å, *Acta Crystallogr D. 59*, 73-76.
3. Caradoc-Davies, T. T., Cutfield, S. M., Lamont, I. L., and Cutfield, J. F. (2004) Crystal structures of *Escherichia coli* uridine phosphorylase in two native and three complexed forms reveal basis of substrate specificity, induced conformational changes and influence of potassium, *J Mol Biol 337*, 337-354.
4. Morgunova, E., Mikhailov, A. M., Popov, A. N., Blagova, E. V., Smirnova, E. A., Vainshtein, B. K., Mao, C., Armstrong Sh, R., Ealick, S. E., Komissarov, A. A., and et al. (1995) Atomic structure at 2.5 Å resolution of uridine phosphorylase from *E. coli* as refined in the monoclinic crystal lattice, *FEBS Lett. 367*, 183-187.
5. Dontsova, M. V., Gabdoulkhakov, A. G., Molchan, O. K., Lashkov, A. A., Garber, M. B., Mironov, A. S., Zhukhlistova, N. E., Morgunova, E. Y., Voelter, W., Betzel, C., Zhang, Y., Ealick, S. E., and Mikhailov, A. M. (2005) Preliminary investigation of the three-dimensional structure of *Salmonella typhimurium* uridine phosphorylase in the crystalline state, *Acta Crystallogr. F 61*, 337-340.
6. Pugmire, M. J., and Ealick, S. E. (2002) Structural analyses reveal two distinct

families of nucleoside phosphorylases, *Biochem. J. 361*, 1-25.

7. Ealick, S. E., Rule, S. A., Carter, D. C., Greenhough, T. J., Babu, Y. S., Cook, W. J., Habash, J., Helliwell, J. R., Stoeckler, J. D., Parks, R. E., Jr., and et al. (1990) Three-dimensional structure of human erythrocytic purine nucleoside phosphorylase at 3.2 Å resolution, *J. Biol. Chem. 265*, 1812-1820.

8. Mao, C., Cook, W. J., Zhou, M., Federov, A. A., Almo, S. C., and Ealick, S. E. (1998) Calf spleen purine nucleoside phosphorylase complexed with substrates and substrate analogues, *Biochemistry 37*, 7135-7146.

9. Mao, C., Cook, W. J., Zhou, M., Koszalka, G. W., Krenitsky, T. A., and Ealick, S. E. (1997) The crystal structure of *Escherichia coli* purine nucleoside phosphorylase: a comparison with the human enzyme reveals a conserved topology, *Structure 5*, 1373-1383.

10. Appleby, T. C., Erion, M. D., and Ealick, S. E. (1999) The structure of human 5'-deoxy-5´-methylthioadenosine phosphorylase at 1.7 Å resolution provides insights into substrate binding and catalysis, *Structure 7*, 629-641.

11. Lee, J. E., Cornell, K. A., Riscoe, M. K., and Howell, P. L. (2001) Structure of *E. coli* 5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase reveals similarity to the purine nucleoside phosphorylases, *Structure 9*, 941-953.

12. Zhang, Y., Cottet, S. E., and Ealick, S. E. (2004) Structure of *Escherichia coli* AMP nucleosidase reveals similarity to nucleoside phosphorylases, *Structure 12*, 1383-1394.

13. Johansson, M. (2003) Identification of a novel human uridine phosphorylase, *Biochem. Biophys. Res. Commun. 307*, 41-46.

14. Roosild, T. P., Castronovo, S., Fabbiani, M., and Pizzorno, G. (2009)

Implications of the structure of human uridine phosphorylase 1 on the development of novel inhibitors for improving the therapeutic window of fluoropyrimidine chemotherapy, *BMC Struct. Bio.l 9*, 14.

15. Kline, P. C., and Schramm, V. L. (1993) Purine nucleoside phosphorylase. Catalytic mechanism and transition-state analysis of the arsenolysis reaction, *Biochemistry 32*, 13212-13219.

16. Miles, R. W., Tyler, P. C., Furneaux, R. H., Bagdassarian, C. K., and Schramm, V. L. (1998) One-third-the-sites transition-state inhibitors for purine nucleoside phosphorylase, *Biochemistry 37*, 8615-8621.

17. Shi, W., Li, C. M., Tyler, P. C., Furneaux, R. H., Cahill, S. M., Girvin, M. E., Grubmeyer, C., Schramm, V. L., and Almo, S. C. (1999) The 2.0 Å structure of malarial purine phosphoribosyltransferase in complex with a transition-state analogue inhibitor, *Biochemistry 38*, 9872-9880.

18. Smith, M. B., and March, J. (2001) *March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*, 5th Edition ed.

19. Gibbs, R. A. (1998) Prenyl transfer and the enzymes of terpenoid and steroid biosynthesis, in *Comprehensive Biological Catalysis* (Sinnott, M., Ed.), pp 31-118, Academic Press, San Diego.

20. Battersby, A. R., and Leeper, F. J. (1990) Biosynthesis of the pigments of life: mechanistic studies on the conversion of porphobilinogen to uroporphyrinogen III, *Chemical Reviews 90*, 1261-1274.

21. Davies, G., Sinnott, M. L., and Withers, S. G. (1998) Glycosyl transfer, in *Comprehensive Biological Catalysis* (Sinnott, M., Ed.), pp 119-207, Academic Press, San Diego.

22. Peapus, D. H., Chiu, H.-J., Campobasso, N., Reddick, J. J., Begley, T. P., and Ealick, S. E. (2001) Structural characterization of the enzyme-substrate, enzyme-intermediate, and enzyme-product complexes of thiamin phosphate synthase, *Biochemistry 40*, 10103-10114.

23. Hanes, J. W., Ealick, S. E., and Begley, T. P. (2007) Thiamin phosphate synthase: the rate of pyrimidine carbocation formation, *J. Am. Chem. Soc. 129*, 4860-4861.

24. Liu, Q., Kriksunov, I. A., Jiang, H., Graeff, R., Lin, H., Lee, H. C., and Hao, Q. (2008) Covalent and noncovalent intermediates of an NAD utilizing enzyme, human CD38, *Chem. Biol. 15*, 1068-1078.

25. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.

26. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D 54*, 905-921.

27. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for the building of protein models in electron density maps and the location of errors in these models., *Acta Crystallogr. A 47*, 110-119.

28. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283-291.

29. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with

SHELXD, *Acta Crystallogr. D 58*, 1772-1779.

30. Collaborative Computational Project-Number 4. (1994) The CCP-4 suite: programs for protein crystallography, *Acta. Crystallogr. D 50*, 760-763.

31. Brodersen, D. E., De La Fortelle, E., Vonrhein, C., Bricogne, G., Nyborg, J., and Kjeldgaard, M. (2000) Applications of single-wavelength anomalous dispersion at high and atomic resolution, *Acta Crystallogr. D56*, 431-441.

32. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D 60*, 2126-2132.

33. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D 54*, 905-921.

34. Kleywegt, G. J., and Jones, T. A. (1996) xdlMAPMAN and xdlDATAMAN-programs for reformatting, analysis, and manipulation of biomacromolecular electron-density maps and reflection datasets, *Acta Crystallogr. D 52*, 826-828.

35. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination, *Acta Crystallogr. D 58*, 1948-1954.

36. van Aalten, D. M., Bywater, R., Findlay, J. B., Hendlich, M., Hooft, R. W., and Vriend, G. (1996) PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules, *J. Comput. Aided Mol. Des. 10*, 255-262.

37. Philo, J. S. (2006) Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques, *Anal. Biochem. 354*, 238-246.

38. Philo, J. S. (2000) A method for directly fitting the time derivative of sedimentation velocity data and an alternative algorithm for calculating sedimentation coefficient distribution functions, *Anal. Biochem. 279*, 151-163.

39. DeLano, W. L. (2002), The PyMOL Molecular Graphics Systems, DeLano Scientific, San Carlos, CA.

40. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol. 233*, 123-138.

41. Smar, M., Short, S. A., and Wolfenden, R. (1991) Lyase activity of nucleoside 2-deoxyribosyltransferase: transient generation of ribal and its use in the synthesis of 2′-deoxyribonucleosides, *Biochemistry 30*, 7908-7912

42. Morgan, P. M., Sala, R. F., Tanner, M, E. (1997) Eliminations in the reactions catalyzed by UDP-*N*-acetylglucosamine 2-epimerase, *J. Am. Chem. Soc. 119*, 10369-10277

43. Lee, S. S., Yu, S., and Withers, S. G. (2003) Detailed dissection of a new mechanism for glycoside cleavage: a-1,4-glucan lyase, *Biochemistry 42*, 13081-13090.

44. Murkin, A. S., Chou, W. K., Wakarchuk, W. W., and Tanner, M. E. (2004), Identification and mechanism of a bacterial hydrolyzing UDP-*N*-acetylglucosamine 2-epimerase. *Biochemistry 43*, 14290–14298.

45. Legler, G. (1990) Glycoside hydrolases: mechanistic information from studies with reversible and irreversible inhibitors, *Adv. Carbohyd. Chem. Biochem. 48*,

319–384.

46. Burmeister, W. P., Henrissat, B., Bosso, C.., Cusack, S., and Ruigrok, R. W. H. (1993) Influenza B virus neuraminidase can synthesize its own inhibitor, *Structure 1*, 19–26.

CHAPTER 5

SUMMARY AND CONCLUSIONS

Over the past years, macro-molecular X-ray crystallography has become a keystone in structural genomics, proteomics, bioinformatics, protein engineering and rational pharmacology. In the presented work, macromolecular X-ray crystallography has been used as a tool for structural characterization of enzymes involved in thiamin biosynthetic and pyrimidine salvage pathways, in eukaryotes. The structural information obtained at atomic resolution level elucidates the overall architecture along with the active site composition of a particular enzyme of interest. The three dimensional structure of the enzyme with substrate, intermediate and product analogue, help in understanding the structural changes occurring along the reaction coordinate providing an insight to the mechanism of catalysis. The structural studies of enzymes involved in various metabolic pathways are relevant because they can be considered as suitable drug targets.

The enzymes involved in thiamin biosynthesis have been a focus of research for the last decade or so (*1*). The structural and biochemical studies have revealed most of the details of this particular pathway in bacteria. The thiazole and pyrimidine fragments are synthesized in two separate routes and then combined together to form ThMP. A final phosphorylation of ThMP leads to the formation of TPP, the active biological form of the cofactor. Although considerable progress has been made in the prokaryotic system, there are still some important questions to be answered. The structure of ThiH and its interaction with ThiG is still not known. The finding that some organisms lack the enzymes needed in the standard pathway, opens up the possibility of existence of some new pathways.

The thiamin biosynthesis in eukaryotes is still in its early stage of development and has a lot of unraveled questions. The recent discovery of the structure of the

eukaryotic thiazole synthase ScTHI4 with the mechanistic elucidation have illustrated that the process is very different between bacteria and eukaryotic systems (*2, 3*). The single enzyme converts NAD, glycine and an unknown sulfur source to an adenylated thiazole carboxylate (ADT). Current experimental details have indicated that the sulfur

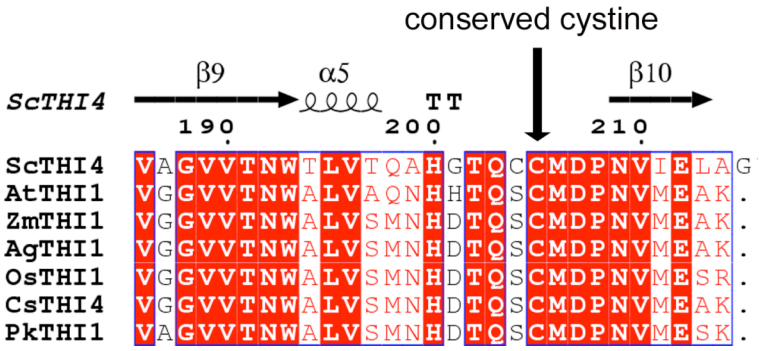


Figure 5.1. Sequence alignment of THI4/THI1 orthologues showing the conserved Cystine residue.

comes from a conserved cystine residue (Cys 205 in ScTHI4) of the enzyme (Figure 5.1). This shows an example where the enzyme acts as a cosubstrate (Abhishek Chatterjee, private communication). However the role of THI4 in mitochondrial DNA repair is still not understood. The pyrimidine biosynthesis in eukaryotes is under development. THI5 has been identified as the working enzyme that generates the HMP moiety, utilizing pyridoxine and histidine (*4*). Further research on thiamin biosynthetic and transport proteins will continue to reveal the remarkable chemical details of this essential cofactor.

In Chapter 2, structural characterization of eukaryotic thiamin biosynthetic enzyme CgTHI6 has been described. This is the first three dimensional structure of a bifunctional enzyme where the N-terminal domain is responsible for ligating the thiazole and pyrimidine fragments and the C-terminal domain has a salvage activity, responsible for an ATP mediated phosphorylation of ThZ-OH to ThZ-P. The enzyme exists as a homohexamer forming a prolate spheroid structure where the two C-

terminal domains form the top and bottom core of the structure, bridged by two N-terminal domains. The structure is also interestingly characterized by a central cavity accounts for the high solvent content of the prepared crystals. Our structural studies reveal that the folds of the domains are very similar to the bacterial counter-parts along with the retention of the trimeric C-terminal and dimeric N-terminal domains as observed in case of corresponding bacterial enzymes, which provides some implication toward the role of evolution in the structure-function relationship.

While the three dimensional structure and the active sites of each domain of CgTHI6 have been identified the possible coordination between the domains is still not clear. The product of the C-terminal domain (ThZ-P) acting as a substrate for the N-terminal domain favors the formation of this bifunctinal enzyme, however there was no obvious channel found between those two domains of a protomer suggesting no possible substrate channeling. In recently studied bifunctional enzymes the demonstration of substrate channeling have been a common phenomena (*5-7*) whereas in this case the substrate ThZ-P, being not labile, is transferred by diffusion from one active site to the other and the enhancement its local concentration is simply beneficial for the chemical reaction.

The other important question raised from our current study is about the conversion of ADT (THI4 product) to CO_2-Thz-P (THI6 substrate). The absence of the adenylyl cleavage activity in CgTHI6 opens up the possibility of existence of an adenylyl transferase enzyme in this pathway. The identification of the responsible enzyme will reveal the biochemical foundation of this essential process.

Chapter 3 includes the structural and biochemical characterization of a thiazole-thiamin biosynthetic enzyme ZmTHI1. Recent studies showed that a mutation in the thiazole biosynthetic gene *thi1_2* results a phenotype called Bladekiller1 (*blk1*) exhibiting premature termination of influorescence and axillary

meristems and poor development of leaf blades. The observed mutation involves the change of Val211 to Met211. The three dimensional structure of the wild type enzyme has revealed the role of Val211 in structural and biochemical features of the enzyme. Val211 although not involved in any direct interaction with the ligand, is located in a short helix, preceding the active site loop containing residues 215-221. A mutation of this residue could possibly result the change in the loop conformation affecting the ligand binding. The results of the biochemical assays further showed that the wild type enzyme purifies with the adenylated metabolites and the mutant does not. The DLS studies have indicated that the wild type enzyme exists as an octamer unlike the mutation suggesting that the octamer formation is also hindered by the current mutation. The mutants designed have also supported the theory that the thiazole synthase activity and octamer formation depends on the steric effect of the side chain. However the crystal structure of the mutants will further prove our current theory.

Intriguingly Cys 220, a conserved residue in the active site loop did not show any electron density for the sulfur atom. This indicates the possibility of involvement of the enzyme as a co-substrate where Cys 220 acts as the sulfur source of the thiazole ring. Although over the past few years a lot progress has been made in the eukaryotic thiazole synthase there are few interesting questions remaining. The role of metal in the activity have been suggested for ScTHI4, however the active site shows no electron density for metal. This observation could be explained by the potential release of the metal ion after the catalysis and product formation.

The second pathway, presented in this work is the salvage of uridine. The enzymes involved in salvage pathways, providing alternatives to the energy costly *de novo* biosynthetic pathways are potential drug targets (*8, 9*). The current work includes structural characterization of UP, that catalyzes the reversible phosphorolysis

of uridine to uracil and ribose 1-phosphate (or 2′-deoxyuridine to 2′-deoxyribose 1-phosphate). The crystal structures with different substrates show the formation of a glycal in the enzyme active site. NMR time course studies elucidate that this enzyme can catalyze the hydrolysis of fluorinated substrates in absence of phosphate or sulfate, without the release of intermediate or enzyme inactivation. The results add a previously unencountered motif to the body of information on glycal formation in active sites of the enzymes, catalyzing the cleavage of glycosyl bonds.

Each of the above examples has touches on active arenas of structural biology results. The fusion of a biosynthetic enzyme and a salvage enzyme by Nature raises the important question of the role of evolution therein in the formation of THI6 in eukaryotes. The formation of blk1 in maize caused by Val211Met mutation illuminates the role of the absolutely conserved amino acid residues in structure-function interplay. Lastly the trapping of glycal in UP active site shows that subtle and still poorly understood effects, occurring during the crystallization process can result in the trapping of intermediates yielding structures rich in mechanistic information. In brief, the structural characterization of above three systems illustrates the contribution of macromolecular X-ray crystallography in the field of structural biology.

REFERENCES

1. Jurgenson, C. T., Begley, T. P., and Ealick, S. E. (2009) The structural and biochemical foundations of thiamin biosynthesis, *Annu. Rev. Biochem. 78*, 569-603.
2. Chatterjee, A., Jurgenson, C. T., Schroeder, F. C., Ealick, S. E., and Begley, T. P. (2006) Thiamin biosynthesis in eukaryotes: characterization of the enzyme-bound product of thiazole synthase from Saccharomyces cerevisiae and its implications in thiazole biosynthesis, *J Am Chem Soc 128*, 7158-7159.
3. Jurgenson, C. T., Chatterjee, A., Begley, T. P., and Ealick, S. E. (2006) Structural insights into the function of the thiamin biosynthetic enzyme Thi4 from Saccharomyces cerevisiae, *Biochemistry 45*, 11061-11070.
4. Wightman, R., and Meacock, P. A. (2003) The THI5 gene family of Saccharomyces cerevisiae: distribution of homologues among the hemiascomycetes and functional redundancy in the aerobic biosynthesis of thiamin from pyridoxine, *Microbiology 149*, 1447-1460.
5. Huang, X., Holden, H. M., and Raushel, F. M. (2001) Channeling of substrates and intermediates in enzyme-catalyzed reactions, *Annu Rev Biochem 70*, 149-180.
6. Miles, E. W., Rhee, S., and Davies, D. R. (1999) The molecular basis of substrate channeling, *J Biol Chem 274*, 12193-12196.
7. Srivastava, D., Schuermann, J. P., White, T. A., Krishnan, N., Sanyal, N., Hura, G. L., Tan, A., Henzl, M. T., Becker, D. F., and Tanner, J. J. Crystal structure of the bifunctional proline utilization A flavoenzyme from Bradyrhizobium japonicum, *Proc Natl Acad Sci U S A 107*, 2878-2883.
8. Al Safarjalani, O. N., Zhou, X. J., Naguib, F. N., Shi, J., Schinazi, R. F., and el Kouni, M. H. Enhancement of the bioavailability of oral uridine by

coadministration of 5-(phenylthio)acyclouridine, a uridine phosphorylase inhibitor: implications for uridine rescue regimens in chemotherapy, *Cancer Chemother Pharmacol. 48*, 389-397.

9. Ashour, O. M., Naguib, F. N., and el Kouni, M. H. 5-(m-Benzyloxybenzyl)barbituric acid acyclonucleoside, a uridine phosphorylase inhibitor, and 2',3',5'-tri-O-acetyluridine, a prodrug of uridine, as modulators of plasma uridine concentration. Implications for chemotherapy, *Biochem. Pharmacol. 51*, 1601-1611.